

Psychological Bulletin

EDITED BY
LYLE H. LANIER, NEW YORK UNIVERSITY

WITH THE CO-OPERATION OF

H. BRITT, McCANN-PACKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HERON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTHWESTERN UNIVERSITY; D. G. MARQUIS, UNIVERSITY OF MICHIGAN; A. W. MELTON, OHIO STATE UNIVERSITY; J. T. METCALF, UNIVERSITY OF VERMONT.

CONTENTS

General Reviews and Summaries:

The Physiological Basis of Visual Acuity: VIRGINIA L. SENDERS, 465.

An Analysis of the Use of the Interruption-Technique in Experimental Studies of "Repression": ALFRED F. GLIXMAN, 491.

The Technic of Homogeneous Tests Compared with Some Aspects of "Scale Analysis" and Factor Analysis: JANE LOEVINGER, 507.

Notes:

Subject and Object Sampling—A Note: KENNETH R. HAMMOND, 530.

Absolute Pitch—A Reply to Bachem: D. M. NEU, 534.

Reply to Postman: G. RAYMOND STONE, 536.

Book Reviews: 538.

Index of Subjects, Volume 45: 551.

Index of Authors, Volume 45: 553.

PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1315 Massachusetts Ave., N.W., Washington 5, D.C.

Subscription price, \$7.00 per year, single issue, \$1.25.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage provided for in Section 1103, act of February 26, 1973, authorized August 6, 1967.

PUBLICATIONS OF
The American Psychological Association, Inc.

AMERICAN PSYCHOLOGIST

Editor: DAHL WOLFE, American Psychological Association

Contains all official papers of the Association and articles concerning psychology as a profession; monthly.
Subscription: \$7.00 (Foreign \$7.50). Single copies, \$.75.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

Editor: GORDON W. ALLPORT, Harvard University

Contains original contributions in the field of abnormal and social psychology, reviews, and case reports; quarterly.
Subscription: \$5.00 (Foreign \$5.25). Single copies, \$1.50.

JOURNAL OF APPLIED PSYCHOLOGY

Editor: DONALD G. PATTERSON, University of Minnesota

Contains material covering applications of psychology to business, industry, education, etc.; bi-monthly.
Subscription: \$6.00 (Foreign \$6.50). Single copies, \$1.25.

JOURNAL OF COMPARATIVE AND PHYSIOLOGICAL PSYCHOLOGY

Editor: CALVIN P. STONE, Stanford University

Contains original contributions in the field of comparative and physiological psychology; bi-monthly.
Subscription: \$7.00 (Foreign \$7.50). Single copies, \$1.25.

JOURNAL OF CONSULTING PSYCHOLOGY

Editor: LAURANCE E. SHAFFER, Teachers College, Columbia University

Contains articles in the field of clinical and consulting psychology, counseling and guidance; bi-monthly.
Subscription: \$5.00 (Foreign \$5.50). Single copies, \$1.00.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Editor: FRANCIS W. IRWIN, University of Pennsylvania

Contains original contributions of an experimental character; bi-monthly.
Subscription: \$7.00 (Foreign \$7.25). Single copies, \$1.25.

PSYCHOLOGICAL ABSTRACTS

Editor: C. M. LOUTHY, University of Illinois, Urbana, Illinois

Contains noncritical abstracts of the world's literature in psychology and related subjects; monthly.
Subscription: \$7.00 (Foreign \$7.25). Single copies, \$.75.

PSYCHOLOGICAL BULLETIN

Editor: LYLE H. LANIER, New York University

Contains critical reviews of books and articles and critical and analytical summaries of psychological fields or subject matter; bi-monthly.
Subscription: \$7.00 (Foreign \$7.25). Single copies, \$1.25.

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Editor: HERBERT S. CONRAD, U. S. Office of Education

Contains longer researches and laboratory studies which appear as units; published at irregular intervals at a cost to author of about \$2.50 a page; author receives 150 copies gratis.
Subscription: \$6.00 per volume of about 350 pages (Foreign \$6.50). Single copies, price varies according to size.

PSYCHOLOGICAL REVIEW

Editor: CARROLL C. PRATT, Princeton University

Contains original contributions of a theoretical nature; bi-monthly.
Subscription: \$5.50 (Foreign \$5.75). Single copies, \$1.00.

Subscriptions are payable in advance and are terminated at expiration.
Make checks payable to the American Psychological Association, Inc.

Subscriptions, orders, and other business communications should be sent to:

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVENUE N.W., WASHINGTON 3, D.C.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN

Psychological Bulletin

THE PHYSIOLOGICAL BASIS OF VISUAL ACUITY

VIRGINIA L. SENDERS

*Wellesley College**

INTRODUCTION

What are the linear dimensions of the smallest visual stimulus to which an organism can react? What are the factors which limit the smallness of the linear dimensions of this stimulus? And what is the physiological basis of the organism's reaction? These are the three main questions which have concerned investigators in the field of visual acuity.

Visual acuity has been defined as the *reciprocal of the minimum visible angle measured in minutes of arc*. The definition is in terms of angles because of the assumption, implicit in the literature, that visual acuity is independent of distance. It has been assumed that the invariant in the case is the visual angle and not the linear measurement of the stimulus. Actually, there is some experimental evidence which suggests that acuity as defined above may depend on distance (3, 18); however, the results are not entirely clear-cut. If it can be shown that acuity is a systematic function of distance, and hence that visual angle is not invariant, the definition will have to be revised. Most modern investigators, however, recognize that there may be a problem here; and although they define acuity in angular terms, they try in their experiments to keep distance constant within small limits. For convenience we shall accept the definition in terms of angle, remembering, however, that the angle has not been proved beyond question to be invariant.

Throughout the literature the term *visual acuity* has been used whenever any sort of minimum visible angle has been under discussion. It has

* Part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Harvard University. The author wishes to express appreciation to Professor William J. Crozier of the Laboratory of General Physiology, Department of Biology, for his assistance in the preparation of this survey, and to Professor Edwin G. Boring, Department of Psychology, for his criticisms of the manuscript.

been used for the minimum discriminable separation of bright points, dark points, lines, the limbs of a broken circle, or the offset of a broken line or contour; it has been used for the minimum angular width of a single hair-line which can just be perceived, and for the complicated stimuli of a Snellen chart. The difficulty with this all-inclusive use of a single term is that when we come to consider the physiological bases of visual acuity, we may tend to assume that since all these sorts of discrimination have been subsumed under the single concept, *visual acuity*, they must all have the same physiological basis.

In order to avoid this difficulty, several authors have suggested subdivisions of the concept. Not all authors agree on what these subdivisions should be. Berger (7), for example, maintains that the measure of acuity, as it has been defined, is dependent on so many physiological and psychological factors—such as the nature of the background or surround, the size of the pupil, the amount of contrast, attention, and the intensity of the stimulating illumination—that he proposes the use of the term *visual resolution* to be applied only to what he considers the simplest case: the minimum separation of two bright points on a dark field. The functional relation between acuity and intensity in this case is different from all other cases; this is a further reason for considering the minimum separation of two bright points a distinct process.

Lythgoë (34) has suggested that it would be desirable to distinguish between the *minimum separabile*, the *minimum visibile*, and the *form sense*, this latter being applied to the perception of complicated shapes. The *minimum separabile* refers to the minimum angular separation required for two points or lines to be perceived as separate; while the *minimum visibile* refers to the minimum angular width required to see a fine line. Since below a certain point, longer lines are more visible than shorter ones, the test line should be just long enough so that length is not a relevant parameter.

Whether or not we make a distinction between the *minimum visibile* and the *minimum separabile*, or between *visual acuity* and *visual resolution* is not particularly important; it is important, however, that we bear in mind that the only thing which these different types of acuity may have in common is that they are all measures of some sort of minimum visual angle. The physiological processes which underlie these different types of discrimination may ultimately be shown to be quite different.

It is important to know the maximum visual acuity which can be obtained under optimal conditions, because until we know what the absolute limit of discrimination is, we cannot determine the anatomical

or physiological factors which limit it. For many years, for example, investigators did most of their experiments on Landolt rings and linear grids. They found that the minimum separation of the bars of a grid was about one minute of arc; anatomists at that time also estimated that the subtense of a foveal cone was about one minute of arc, and so it was assumed that the limiting factor in visual acuity was the size of the foveal cone. We know today that measures of much higher acuity may be obtained with other kinds of test objects, and we have also revised our estimate of the size of the foveal cones so we can no longer accept cone-size as the limiting factor. However, our manner of approach must remain the same; we must determine from our knowledge of the structure of the organism, and from our knowledge of the organism's reaction, the anatomical and physiological mechanisms which underlie and limit that reaction.

EMPIRICAL ISSUES

The main theoretical problem in the field of visual acuity is to determine the nature of the physiological processes underlying those various types of discriminations which have been subsumed under the concept *visual acuity*. But before we can consider theories, we must know certain empirical facts. We must know the minimum visible angle which has been obtained under each of several conditions; most investigators ask further: how does this angular size compare with the size and separation of receptors in the retina? From a comparison of these two figures, what conclusions is it possible to draw regarding the retinal basis of visual acuity? The second group of empirical facts, often studied in conjunction with the first, concerns the nature of the functional relation between visual acuity and intensity of illumination. A brief summary will be presented of the major investigations and results on each of these problems.

Determinations of the Minimum Visible or Separable Angle

As might be expected, there is considerable variation in the estimates of the minimum visible angle. Much of this variation appears to be due to differences in the nature of the target used. The results of the investigations will therefore be classified according to the type of target or test-object. Unless otherwise stated, all the determinations to be reported have been made in the fovea.

Bright points on a dark field. The classical observation on the minimum angular separation of two bright points which are to be seen as separate was made by Helmholtz ((31), who reported this separation to be 1' of arc.

Because of eye-movements and diffraction in the ocular media, the retinal image of a bright point is many times larger than the simple geometrical image would be. It might be expected, then, that because of the increased size of the blur-circle due to an increase in intensity the minimum separation of two bright points would be found to *increase* rather than to decrease with an increase in intensity. As intensity increases, visual acuity should decrease. This is exactly what was found by Berger (7) and confirmed by Berger and MacFarland (8). Berger obtained the surprisingly large angular separation of 180''–200'' when the measurement was made at the absolute threshold for brightness.

The separation found by Berger was more than double that later found by Tonner (41). This author found a minimum separation of approximately 73''. He attempted to control pupil diameter by stimulating the non-observing eye with fairly high light intensities, thus constricting the pupils of both eyes. This method of controlling pupil size may be criticized, since the effects on one eye of stimulation of the other are not yet fully understood (*cf.* 23, 33) but even this does not explain the large discrepancy between these two sets of results.

Single dark line. Some of the early investigations were concerned with the visibility of a single line. In some cases wires, spiders' webs, or hairs were viewed against white paper or against the sky. Some of these values are cited by Hartridge (24) and others by Walls (44). The values obtained varied between 0.44'' and 6'' of arc. They were:

Aubert	6''
Smith and Kastner	3.5''
Hartridge	3.6''
Hartridge and Owen	3.1''
Lowell	0.83''
Pickering	0.85''
Barnard	0.44''

In all of these cases the angular length was quite large, usually covering at least several minutes. The last three experimenters all used the sky as a background, and possibly the use of longer fixation distances, or longer angular lengths enabled them to obtain estimates lower than those of the other investigators.

Hecht and Mintz repeated these earlier investigations under carefully controlled laboratory conditions (26). A single hair-line against a very evenly illuminated background was used by them because they considered it to be the simplest case of visual resolution. The observer was dark-adapted for twenty minutes, a natural pupil was used, and the experimenters attempted to minimize the influence of distance by keeping the observer at a distance of from two to three meters from the target. An indefinite fixation time was permitted. The line was long enough so that length was no longer a critical factor.

They found that the minimum width of the line varied from 10' to

0.5'' of arc, depending on the illumination of the background. Their limiting angle of 0.5'' is probably the best estimate available at present of the absolute *minimum visibile*.

Distance between bars of a grid. Although the procedures used by various experimenters differed in such details as the breadth of the bars and the use of monocular or binocular regard, the values obtained are quite similar. Several are cited by Hartridge (24):

Lister	64''
Hirschmann	50''
Bergmann	52''
Helmholtz	64''
Uhtoff	56''
Cobb	64''

Since these angles are about the same as the earlier estimates of cone-width, the conclusion was drawn that the limiting factor in visual resolution was cone-size; two points or lines could be perceived as separate if each one stimulated a single cone on either side of a non-stimulated one. However, as has been seen, the measures obtained with other types of test-objects do not confirm this hypothesis, nor do more recent estimates of cone-size.

Vernier acuity. Vernier acuity may be defined as the minimum lateral displacement necessary for two portions of a line to be perceived as discontinuous. Polyak (36) cites values for vernier acuity as low as 2.5'' of arc but does not give his sources; Titchener cites Hering as having obtained a value of 5'' of arc (40) and some values reported by Hartridge (24) are:

Bryan and Baker	Black lines	12''
Bryan and Baker	White lines	9.5''
Bryan and Baker	Split lines	8''
Bryan and Baker	Bisection lines	8.5''
Hartridge	Black lines	8.5''
Stratton	Black lines	7''

Stereoscopic acuity. Stereoscopic acuity may be defined as the just perceptible difference in binocular parallax of two objects or points. The minimum values obtained for stereoscopic acuity are quite similar to those obtained for vernier acuity. Some which have been reported by Hartridge (24) are:

Pulfrich	10''
Heine	6''-13''
Bowdon	5''
Crawley	3'' (approximately)
Breton	4''

Stereoscopic acuity and vernier acuity are similar in that both involve more complicated perceptual processes than the separation of bars or

points or the visibility of single lines. It has been suggested that the retinal and neural processes involved in both stereoscopic and vernier acuity may be different from those of the more usual type of target.

Summary. In comparing the results recorded by different investigators in an attempt to determine, under the best possible conditions of adaptation, illumination, and so forth, the maximum visual acuity, we find that the values of the minimum resolvable angle range from $0.5''$ to $200''$. This is a range of 1 : 400. It will be noted that for a particular type of test-object the measurements tend to be far more homogeneous than do measurements for a variety of test-objects; however, there are other possibilities for variation, due to experimental conditions other than the mere type of test-object used. The measure of visual acuity has been found to depend on pupil-size, on the size and brightness of the surround, on wave-length, on area, on the use of monocular and binocular regard, and possibly on distance.

The best resolution is for a single line, followed by that for stereoscopic and for vernier acuity. The distance between the bright bars of a grid is considerably larger than the minimum angles obtained for lines, offsets, or parallax, and (except in the case of Berger's results) is only slightly less than the angles obtained for the separation of two bright points.

We now have some good estimates of the maximum human capacity for visual spatial discrimination under different experimental conditions. It is natural then to ask about the structure of the mechanism which makes this discrimination possible. Since it has repeatedly been suggested that the size of the foveal cones is the limiting factor in visual resolution, some estimates of cone size will be presented, and their relation to the obtained measures of minimum visual angle will be discussed.

Comparison with Cone Size

The diameter of a foveal cone has been variously estimated at between 1.0μ and 5.4μ . Some of these values are cited by Clemmesen (11):

Henle	3.0μ
Kolliker	3.0μ
Schultze	2.5μ
Koster	4.5μ
Greeff	2.5μ
Heine	4.0μ
Rochon-Duvigneaud	$2.0\mu-2.2\mu$
Fritsch	$1.8\mu-4.5\mu$

Polyak (36), however, makes much lower estimates. Cones in different portions of the fovea are of different sizes, the smallest being in the inner

fovea or *foveola*. Here he estimates that the average diameter of a cone is 1.0μ at its tip and 1.3μ at its base. The diameter of the cones increases as we go into the outer fovea up to an estimated 3.5μ to 4.0μ . In any such estimates of cone size there remains the difficulty of extrapolating knowledge gained from a fixed preparation to the state of affairs in the living organism.

Throughout the literature, the minimum resolvable separation of two bars, which is of the order of magnitude of one minute of arc, has been cited in proof of the assumption that the *minimum separabile* is limited by cone diameter. According to the early estimates, which were undoubtedly too large, one cone diameter subtended an angle of about one minute. It was therefore assumed that there had to be one unstimulated cone between two stimulated ones in order for a separation to be perceived. However, if we accept Polyak's estimate, and assume the focal length of the eye to be 15 mm., then one cone subtends an angle of $20.6''$, or only one-third of a minute. The separation in the case of bars turns out to require at least three unstimulated cones between the stimulated ones.

On the other hand, Volkmann, in 1892 (43), was the first to point out that the resolving power of the eye may in some cases be a great deal better than one minute of arc, and his observations were confirmed by Wülfing (48) and by Hering (24). Even with the most recent estimates of foveal cone diameters, the minimum values for stereoscopic acuity, vernier acuity, and the visibility of a single line lie well below the diameter of a single cone. Under a variety of conditions the minimum detectable angle may be far too small to be accounted for by the simple "cone separation" theory; lately the effort has been made to establish a theory which will adequately account for the discrepancies between the observed facts and the primitive theory.

The Visual Acuity of Insects

Another sort of investigation which has attempted to determine the relation between visual acuity and the size of the receptor units has been the studies of visual acuity in insects.

In 1929 Hecht and Wolf made an intensive study of the visual acuity of the honey bee (30). They used the nystagmic response of the bee as an indication that the images of moving stripes in the visual field had been resolved. They found that the maximum visual acuity (the reciprocal of the minimum visible angle in minutes of arc) for the honey bee was between 0.016 and 0.017. This is below the visual acuity of the human eye at the lowest perceptible illuminations.

Basing their calculations on Baumgärtner's investigations (6) of

the size and distribution of ommatidia in the bee's eye, they concluded that the visual angle corresponding to the maximum visual acuity was identical with the angular separation of adjacent ommatidia in the region of maximum density of the ommatidial population. Since the ommatidia are not evenly distributed throughout the bee's eye (according to Baumgärtner), they tested this coincidence of values by rendering non-functional the region of the eye with the greatest density of ommatidia. This was done by placing a drop of black paint on the surface of the eye. A decrease in acuity was found. The conclusion was drawn that the minimum visible angle is determined by the separation of adjacent ommatidia; since ommatidia are assumed to have thresholds distributed statistically throughout the entire intensity range in which discrimination is possible, an increase in illumination results in a "functional decrease" in the separation of excitable adjacent elements.

The conclusions drawn by Hecht and Wolf may be criticized on anatomical grounds. The estimates of the size of ommatidia which they used were taken from Baumgärtner, who estimated the minimum angular separation of the centers of adjacent ommatidia to be $51'$ of arc. A more recent study, using better histological techniques (16), indicates that the minimum angular separation is actually $1^{\circ}18'$, or 50% larger than Baumgärtner had estimated. In other parts of the bee's eye, Baumgärtner's estimates are too small by as much as $50'$. Thus the results of Hecht and Wolf would indicate that one and a half ommatidia rather than a single ommatidium, are included in the minimum separable angle.

A second study of the visual acuity of insects was that made by Hecht and Wald in 1933 (29). They investigated the visual acuity of *Drosophila*. The procedure was very similar to that used by Hecht and Wolf. They found that the maximum visual acuity achieved by *Drosophila* was 0.0018, a value about 1/1000 that of the human eye, and 1/10 that of the bee's eye.

The minimum angular separation found by Hecht and Wald was $9^{\circ}17'$. They themselves prepared *Drosophila* eyes and found the minimum ommatidial separation to be $4^{\circ}12'$. Thus if the limiting factor in resolution were the separation of stimulated receptors by non-stimulated ones, the minimum separation would include two ommatidia, even at the highest intensities. This result was attributed to the small number of receptors in the eye of the animal, although the possible role played by neural connections between ommatidia was also considered.

It has by now become apparent that when we consider the empirical findings, both anatomical and experimental, we are unable to account

for maximum visual acuity *solely* in terms of the size and separation of receptor elements. But it has long been known that visual acuity is better under high illuminations than under low; perhaps an explanation which includes some consideration of the dependence of acuity on intensity, as well as on cone diameter and separation, can adequately account for the observed findings.

Visual Acuity as a Function of Intensity

Some of the early investigators who were interested in the problems of visual acuity came to the consideration of the influence of intensity of illumination on the minimum separable or visible angle. The classical experiments were those of Koenig (32) in 1897. His experimental conditions would be considered crude today, since he used test-objects made of black and white paper illuminated from in front, varied the illumination in three different ways, did not control pupil-size or surround, and altered the distance of the test-object from the observer. However, Koenig's general results, which showed that visual acuity, as previously defined, varied in sigmoid fashion with the logarithm of the intensity of the illumination, have been repeatedly verified since. He put three straight lines through his data and developed empirical equations to fit them, but it is his data, and not his explanation, which have become classical.

In more recent years other workers, notably Hecht and his associates, have repeated Koenig's experiments, with some modifications, and have substantially confirmed his general results. Shlaer (37) investigated the dependence of visual angle on illumination for Landolt rings and for linear grids. His results are in fairly good agreement with those of Koenig.

When visual acuity is plotted against the logarithm of the intensity, the function is sigmoid to inspection; but when the *logarithm* of visual acuity is plotted against the logarithm of the intensity, a discontinuity in the function appears at about zero log units of intensity (photons). This discontinuity is attributed to the determinative functioning of rods at low intensities and of the cones at higher intensities.

Shlaer concludes that the two factors which operate to limit resolution are pupil-size and diffraction by the pupil, and the separation of retinal elements. Visual acuity increases with pupil-size until the diameter of the pupil is 2.3 mm, and thereafter remains constant. When pupil-size was not the limiting factor, the maximum visual acuity was 2.1, which Shlaer believes represents a spacing exactly equivalent to that of the retinal receptors.

Shlaer's theoretical interpretation of these visual acuity data is

essentially the same as Hecht's. In the case where the separation of retinal elements is considered to be the limiting factor, he explains the increase in acuity with increased intensity by the functional increase in number of available receptors due to the successive involvement of those with higher and higher thresholds. Further discussion of this theory is deferred to a later point. We must first review Hecht's experiments and then discuss his interpretations.

A general outline of the procedure used by Hecht and Mintz in determining the minimum visible angle of single hair-lines has already been presented. Their results also showed the usual relationship between visual acuity and illumination: visual acuity increases in sigmoid fashion with the logarithm of the stimulating intensity. Plotting the logarithm of visual acuity against the logarithm of intensity, they also found a discontinuity similar to that found by Shlaer.

With a few exceptions (8, 46) visual acuity is seen to increase with an increase in stimulating intensity in a regular and lawful fashion. Such a functional relation must be accounted for by any theory of visual acuity.

THEORIES OF VISUAL ACUITY

Visual Acuity as a Form of Intensity Discrimination by the Retina

Values for vernier and stereoscopic acuity, and for the visibility of single lines are too small to be accounted for by the primitive "cone separation" theory of visual acuity; the values obtained for the separation of bars or points are too large to be accounted for in this way. In view of this fact, and in view of the known functional relation between visual acuity and intensity, it seemed reasonable to many authors to assume that the resolving power of the retina was a function of the distribution of intensities on the retina. Visual acuity has thus come to be considered a particular form of intensity discrimination.

Hartridge's theory. Hartridge (24) in 1922 stated that visual acuity was dependent upon the resolving power of the retina, and that this resolving power was dependent upon intensity discrimination in certain specified ways. The essential point made by Hartridge was that the intensity difference between "stimulated" and "unstimulated" cones need not be 100% as the older theories had suggested, but might be as little as 5%. The image is thought of as a distribution of intensities on the retina, of which even the center, or darkest portion, has some illumination (due to diffraction and aberrations). Thus for a single resolved line, if the illumination of a cone upon which the image does not fall is taken to be 100%, the average intensity upon the central cones of

the image is calculated (from Rayleigh's equations) to be 83%, while the cones on either side of the center are calculated to have an illumination of 96%. This difference of 13% is comparable to the 10% intensity difference which Hartridge believes is necessary for simple intensity discrimination. Similar calculations are made for other types of test-object, and in each case the minimum visible angle is held to be dependent upon intensity discrimination for that particular type of test-object. Thus Hartridge is to be remembered for having formulated, in quantitative terms, a theory which attempts to explain visual acuity as a particular type of intensity discrimination.

One criticism of this theory has been advanced by Byram (10). He points out that Hartridge's calculations of intensity distributions are based on Rayleigh's equations; but Rayleigh's equations were intended for use with a rectangular aperture and are not correct when a circular aperture is used. The pupil of the eye is, of course, not rectangular, and therefore many of Hartridge's calculations are in error. Byram has calculated that some of his values are more than twice as high as they should be.

Photochemical theory. Hecht's explanations of his results, and his theory of visual acuity, are essentially extensions and modifications of Hartridge's earlier theory. The explanation offered by Hecht and Mintz of the small angular width of a single line which could be resolved runs somewhat as follows: the eye is by no means a perfect optical instrument, hence the calculated geometrical image will not be the same as the retinal image. Chromatic and spherical aberration, as well as diffraction by the various media of the eye, combine to produce a retinal image which is considerably more diffuse than the calculated simple geometrical image would be. They therefore point out that what we have is a distribution of intensities on the retina.

Hartridge has estimated that a 5%-10% difference in illumination between adjacent cones is sufficient to be discriminated. (Actually this is a questionable finding, since such discrimination has been shown to be a function of area (39, 42). Hecht considers this estimate too high, and believes that a difference of 0.95% is sufficient. This is a Weber ratio of 1/105. The line appears sharp, rather than fuzzy, he believes, because only one row of cones is illuminated enough to be stimulated.

The capacity for intensity discrimination increases with an increase in illumination, so, according to Hecht, it is reasonable to assume that resolution, which appears to be a form of intensity discrimination, should increase in similar fashion. Experimental results show that for a fixed width of line and variable intensity, the line is resolved only for intensities at which the brightness difference between the line and its surround is also perceptible. This same result holds when the intensity is fixed and the breadth of the line is varied.

Furthermore, if visual acuity is dependent upon intensity discrim-

ination, then the functions for visual angle and intensity, and for $\Delta I/I$ and intensity, should be similar. When the logarithm of the visual angle is plotted against the logarithm of the stimulating intensity, the curve is seen to be of form similar to that of $\Delta I/I$ plotted against $\log I$.

The argument continues, visual acuity may be considered to be the resolving power of the retina. Thus differences in visual acuity must correspond to differences in resolving power. Since resolving power must be a function of the number of elements per unit area, it is fixed anatomically; if it is to vary at all, it must vary functionally. We may assume that the sensibilities of the individual rods and cones are not all the same, but are distributed in the manner of populations. At the lowest illuminations, only a very few units are operative, those with the very lowest thresholds; therefore, the resolving power will be poor and visual acuity will be low. As the intensity increases, more rods or cones will be functional per unit area, and the visual acuity will increase.

A further argument used in support of this theory runs somewhat as follows: there should be a minimum area which will carry out all the functions of the retina. Koenig computed that there were 572 discrete steps of intensity recognition of which 30 could be attributed to the rods and the rest to the cones. Change, according to Hecht's theory, would correspond to the loss or addition of one single receptor unit. The lowest visual acuity value obtained is for 0.03 log units: this is for a visual angle of $44'$, which corresponds to 0.02 mm. on the retina. The minimum retinal area would then be 0.04 sq. mm. There are, Hecht says, 13,500 cones per sq. mm. in the fovea, so the minimum retinal area of 0.04 sq. mm. would contain 540 cones. In the periphery it would contain 60 rods, but since there are multiple connections in the periphery, these figures actually appear at first to exhibit striking agreement with Koenig's classical data.

The distribution of retinal units in terms of populations is fundamental to the photochemical theory. The basis of this distribution is the reversible photochemical system. The sensitivity of a given retinal receptor is assumed to depend on the concentration of decomposition products necessary to discharge an impulse in the attached nerve fiber, and the total number of active elements is a linear function of this concentration. The number of active elements is thus described by the photostationary state equation:

$$KI = \frac{x^2}{a - x}$$

where a is the initial concentration of photosensitive material S , x is the concentration of decomposition products (A and P), I is intensity, and K is a constant.

With specific reference to the form of the visual acuity—intensity function as found in their investigations, Hecht and Mintz develop the following equations:

$$\frac{\Delta I}{I} = c \left[1 + \frac{1}{(KI)^{1/2}} \right]^2$$

where c is the minimum value of $\Delta I/I$ at the highest I , and K is the reciprocal of the intensity at which $\Delta I/I$ is 4 times the minimum. This represents the usual form of the photochemical equation applied to intensity discrimination, and apparently adhered to by their data. But the minimum visible angle may be considered some function of $\Delta I/I$. Therefore

$$\alpha = b' \frac{\Delta I}{I}$$

where α represents visual angle, and b' is a constant. Therefore:

$$\alpha = b \left[1 + \frac{1}{(KI)^{1/2}} \right]^2$$

where $b = b'c$. The constant b fixes the curve on the ordinate, and K fixes it on the abscissa. According to Hecht, this equation provides a satisfactory fit both for his own visual acuity data and for those of others.

To summarize the photochemical theory of visual acuity: the limit of resolution is set by the pattern of units in the receptor mosaic. The greater the number of retinal units, the finer will be the resolution. Since the number of retinal units cannot vary anatomically, it must vary functionally; and this functional variation comes about through changes in the number of individual units active at any given time. The number active depends upon the thresholds of the receptor units, and different receptors are stimulated by differing amounts of photochemical decomposition. The amount of photochemical decomposition depends in a systematic way upon the intensity of the stimulating illumination, hence visual angle is also a systematic function of intensity. *Visual acuity is therefore considered a form of brightness discrimination by the retina.*

Criticisms of the photochemical theory. This theory merits careful analysis. It may certainly be admitted freely that visual acuity, under most conditions, increases with an increase in intensity. It is also undoubtedly true that the size, separation, and number of retinal elements bears some relation to the minimum visual angle. It may, however, be unjustified to draw the conclusion that the entire basis of visual acuity is intensity discrimination by the retina.

There are three main lines of argument against the photochemical theory of acuity. First, we should inquire into the correctness of the empirical facts. If Hecht and his coworkers are basing their assumptions on facts which are incorrect, or which have been shown to be incorrect by later investigations, we should know the effects of this on

their theory. Second, we should know whether their theory is adequate to account for the obtained data. If not, where does the inadequacy lie? Finally, we should know if their interpretation is the best one to account for the obtained data. What other interpretations can be made from the same findings, and are these other interpretations sounder than those made by Hecht?

Are any of Hecht's facts incorrect? It would appear that in a few cases they are. For example, Hecht maintains that the striking agreement between the number of foveal cones, the number of discrete steps of intensity recognition, and the minimum visible angle is evidence for the intensity discrimination theory of visual acuity. He bases this argument on the fact that there are 13,500 cones per square millimeter in the fovea. This is one estimate; there are other estimates, more recent and probably better. One of the best authorities on the anatomy of the retina is Polyak, who estimates (36) that in the inner fovea and the foveola the number of cones per square millimeter is approximately 55,000; in the foveola alone this density is much greater. The minimal retinal area of 0.04 sq. mm. will thus be seen to contain not 540, but 2200 cones, or approximately four times the number calculated by Hecht.

Furthermore, the number of just noticeable differences is a function of area as well as of the intensity level and exposure times used. Had a different size of test-patch, a different intensity level, or a different exposure time been used, the results would have been different. There is no reason to assume that the results obtained by Koenig, with his particular area, exposure time, and intensity level are "standard," or are more valid than those obtained under any other conditions. Thus Hecht is basing some very generalized conclusions on results obtained under one particular set of experimental conditions.

This general line of argument also falls down because it fails to take into account the statistical nature of the *j.n.d.* A *j.n.d.* is defined as that intensity difference which is perceived a given per cent of the time. For any given determination, the required difference may be larger or smaller than this. If a *j.n.d.* represented the loss or addition of a single receptor, as Hecht hypothesizes, then it would have to be admitted that receptor thresholds were variable, rather than constant as Hecht believes. If receptor thresholds were fixed, then the *j.n.d.* would also be a constant quantity, which it is not.

Further, the criticism applied by Byram to Hartridge (10) applies also to Hecht and Mintz. Their calculations of intensity distributions on the retina are based on Rayleigh's equations. But Rayleigh's equations were intended for use with a rectangular aperture, which the eye pupil is not. Therefore, the calculations are too high, although in this case Byram has calculated that the discrepancy is only 15%.

Next we come to the criticism that the photochemical theory is not always *adequate* to account for the obtained results. It offers an explana-

tion, but not always a full explanation. Hecht and Mintz speak of a brightness difference between adjacent cones as the basis for the resolution of a line. However, if no other factors are involved it is difficult to see how the apparent sharpness and straightness of the line can be accounted for, in view of the unequal distribution and spacing of foveal cones. Furthermore, they seem to assume that the distribution of intensities on the retina is essentially static, whereas we know that the image actually "flutters," or shifts back and forth across several cones due to the irregular, saw-tooth movements of the eyes (1). Weymouth, Andersen, and Averill (45), and Andersen and Weymouth (2) have suggested a theory which does not differ in its fundamentals from that of Hecht, but which makes use of both the irregularities of the cone mosaic and the nystagmic movements of the eye. According to these authors we are able to perceive a straight line rather than an irregular or fuzzy one, because the distribution of light intensities shifts back and forth across the retina, and the final perception is the result of retinal "averaging." They do not offer any explanation of how this "averaging" takes place.

The fact that there are eye-movements, and that such movements take a finite amount of time, is one reason why it is desirable to have studies of visual acuity using short exposures. A few such experiments have been performed. Graham and Cook (21) varied intensity, exposure time, and interspace. They used durations from 2 ms. to 500 ms. and found that up to a critical duration, visual acuity improved with time. Averill and Weymouth (4) also used short exposure times. They also found that visual acuity improved with time, but the time-intensity product was *not* constant. The stimulation time of 30 ms. did not require an intensity 50 times greater than that for a stimulation time of 1500 ms. Actually they found that in this case, in which the ratio of the exposure times was 1:50, the ratio of the required intensities was 1:7. These results they attribute to the influence of eye-movements, but other factors, such as the neural recovery cycle, may also be of importance here.

This theory of "retinal averaging" also accounts for the influence of the length of the line on its visibility by demonstrating that the "averaging" process works better for a longer line than for a shorter one. Hecht and Mintz cannot account for the influence of length of line. It is known (4) that the minimum visible angle decreases as the length of line is increased up to a critical length. Hecht and Mintz, however, merely state that their target was longer than the critical length. It is difficult to see how a theory of acuity based only on brightness discrimination could account for such a result unless it included some concept of retinal averaging. It seems very probable that the increase in contour, or edge, is of some importance here (*cf.* 14), but no photochemical explanation can be offered for the influence of contour on visual functions.

A further factor which is not adequately accounted for by Hecht's theory of visual acuity is the effect of adaptation. This effect has been studied by several different investigators, but their conclusions are not entirely concordant. According to the photochemical theory, visual acuity improves with an increase in the size of the population of receptor units; therefore, acuity should be highest with the dark-adapted eye. However, this is apparently not the case. There have been two general methods of studying the effect of adaptation on visual acuity. In one method, the minimum visible angle is determined at intervals during dark-adaptation; in the other method, the observer is adapted to each of several different intensities, and the minimum visible angle for test-objects at various levels of illumination is determined. Experimenters who have employed the first method have generally found that acuity does not improve with dark adaptation as much as does absolute sensitivity. Acuity is usually found to be poor when the observer is dark-adapted, except for test-patches of very low intensities (12). The usual result found when the second general method is employed is that visual acuity is best when the adapting intensity and the intensity of the test-object are the same (17, 34). These results cannot be adequately accounted for by the photochemical theory of visual resolution.

Furthermore, as has been pointed out (9), the variability in organic responses is too great to permit interpretation in terms of retinal units of fixed thresholds. The changes in visual acuity over a range of eight or more log units are too great to be accounted for in this manner. Crozier has also been able to infer from the behavior of the flicker response contour that at any given intensity all the potentially available neural units are participating. Therefore, effects produced at any intensity cannot be accounted for by positing a small number of receptors which are excited because their thresholds have been reached. Furthermore, the assumption that receptors have fixed thresholds has very recently been directly shown to be incorrect, at least for the *Limulus*. Hartline (22) has demonstrated that in this animal the intensity threshold for a single receptor varies by as much as a log unit over a period of time. The total range of stimulus intensity to which the animal can respond is only about four log units.

The theory also falls down because it cannot adequately account for the discrepancies between the functions for a "C," a hair-line, and a grid. The equation derived by Hecht and Mintz is held to fit the results for a "C" and for a "hook," as well as for a hair-line; but it does not fit the data for a grid. This is attributed to the fact that the limiting factor in the resolution of a grid is supposed to be the width of a single cone. (But later estimates of cone-size have shown that this argument is invalid.) Shlaer, Smith, and Chase (38) also find that the same equations are not applicable to both the grid and the "C," and find that, according to these equations, in the case of the "C" the data are adequately fitted by curves drawn to the equation:

$$KI = \frac{x^n}{(a - x)^m}$$

where m and n are the orders of the photochemical and thermal "dark" reactions respectively, and $m = n = 2$. Visual acuity is taken as proportional to x^n , although no reason is given for this. In the case of the grid, when pupil-size is not the limiting factor, the same equation is used, but visual acuity is now taken as proportional to x instead of x^n . This result does not hold when the pupil is the limiting factor. It is difficult—in fact it is impossible—to see how a change in the form of the target, or a change in pupillary diameter, can change the order of a photochemical reaction. It seems more reasonable to look for some sort of effect of the amount of "edge" or perimeter on the visual function. Such an effect was found by Crozier and Wolf (14) in their experiments on flicker with subdivided fields; and, as they point out, it is unreasonable to suppose that when a square field is subdivided by a central cross, the order of the photochemical reaction changes, as would be required to account for the alteration of the parameters of the flicker contour.

However, if we accept Hecht's facts, there may still be some question about how they are to be interpreted. Do his theoretical curves really fit this data? If they do, how much is this due to the use of arbitrary fitting constants? Are his conclusions correct, or could other, more appropriate conclusions be drawn? These questions we shall examine now.

Hecht says that the distribution of sensibilities in the manner of populations is fundamental to his theory of visual acuity. The basis of this distribution is the photostationary state equation. Hecht feels justified in arriving at this conclusion because the photostationary state curves "fit" the obtained data. This question of curve fitting deserves extensive consideration, but only a few words will be devoted to it here. What is the most acceptable criterion of goodness of fit? In general, there are three main classes of criteria: (a) inspection, (b) statistical (*e.g.* least squares) and (c) parametric analysis. Each type of criterion is suitable to some types of data, and in a practical sense may be inapplicable to others.

Essentially the same curve may be obtained from totally different equations. Crozier, for example, has pointed out (15) the complete formal identity of the log logistic and the photostationary state equation as used by Hecht. In many cases, the curves predicted by the photostationary state equation and those predicted by the normal probability integral are so similar as to be indistinguishable by any visual criterion, and a statistical criterion is sometimes also inadequate. In such a case, the most suitable way to distinguish between the curves is by an analysis of the parameters of the function. Crozier has been able in many cases to predict the behavior of the three parameters of the normal probability integral for visual data (the standard deviation, the abscissa of

inflection, and the maximum to which the curve rises) when experimental conditions were systematically altered. This sort of analysis has not been made in terms of the photostationary state equation. When Hecht says that his curves fit his data, he means that a good visual inspection fit is obtained. In the absence of further analysis, this cannot be taken to mean that the photostationary state equation describes the data better than any other.

In deriving the visual angle function from the $\Delta I/I$ function, several constants are involved: c is the minimum value of $\Delta I/I$ for the highest I . K is the reciprocal of the intensity at which $\Delta I/I$ is a minimum. But both of these values are determined by the particular conditions used in a given experiment, and the constants are actually used as arbitrary fitting constants. It has not been possible thus far to test the appropriateness of their use because so many variables are changed at one time (intensity-level, visual angle, distance) that the several effects are not separable.

Furthermore, the fact that the two curves are apparently similar does not necessarily mean that they represent the same basic process. Shaler, it will be remembered, found a break in his curve at about 0 log units of intensity (photons), and this break was attributed to the determinative functioning of rods at low intensities and cones at higher intensities. Since this break occurred at about the same intensity as that found by Hecht, Peskin, and Patt for intensity discrimination (27), this fact was adduced as evidence for the basic identity of intensity discrimination and visual acuity. However, it should be noted that Shlaer used a 30° field, while Hecht, Peskin, and Patt used a 12° field surrounded by a larger, 40° field. In the intensity discrimination experiment, the variable field was exposed to the observer for 1/25 of a second only. Results from experiments on flicker indicate that area may play an important role in determining the location of the break (14). It may safely be predicted that if the two experiments had been conducted under identical conditions, the results would have been different. Thus the appearance of the discontinuity in both curves indicates the functioning of two sets of receptor units, but the fact that it appears at the same intensity cannot be adduced as evidence for the identity of the two functions.

Nor does the similarity between the shapes of the intensity discrimination functions and the visual acuity functions necessarily mean that one is a special form of the other. Wolf (47) was able to show that the $\Delta I/I$ curve could be calculated from the visual acuity measurements, even though in the intensity discrimination measurements the size of the stripes in a rotating cylinder was kept constant while the illumination was changed, while for the visual acuity measurements the intensity was kept constant and the angular size of the stripes was altered by altering the distance of the bee from the stripes. Undoubtedly the visual inspection fit of the derived curves is excellent. However, the

interpretation—that visual acuity is therefore a special case of simple intensity discrimination, is not of necessity the best one. Crozier (13) has shown that $\sigma_{\Delta I}$ is directly proportional to ΔI . Thus by determining from one response contour (either flicker or visual acuity in this case) the values of $\sigma_{\Delta I}$ for different intensities, it is possible to predict the value of ΔI for the corresponding intensity on the intensity discrimination function. The data used by Crozier were actually those obtained by Hecht and Wolf (30) and by Wolf (47). From this agreement, however, he draws several conclusions which are *not* the same as those drawn by Hecht and Wolf.

Two intensities, he says, are not compared, but their *effects* are. The effects of a given intensity are certainly not the same from moment to moment (*cf.* 22); the effects due to the standard intensity are compared with the effects due to the just discriminably greater intensity, and since in both cases these effects are variable, the result of the comparison will not be constant, but will be distributed according to the variabilities of the effects of the two intensities. Thus "it is reasonable to expect that ΔI should be determined by the properties of the two frequency distributions" (13, p. 414).

The essential difference between the viewpoints of Hecht and of Crozier on this point should be noted. Hecht speaks of a population of receptor units, each of which has a fixed threshold. The number of units functioning depends solely upon the intensity and should be constant for any given intensity. *The source of variability is not in the organism but in the stimulus*, as he has attempted to show in his paper on the minimum number of quanta for visual excitation (28). For Crozier, *variability is an important property of the reacting organism*, which is as predictable and as regular as any other function of the organism. Crozier and Wolf, in their series of papers on the flicker response contour, have been able to show that the behavior of σ_I , which is interpreted as a measure of the variability of the organism, may be predicted from changes in the experimental conditions, and that its behavior follows certain known laws.

It appears that the simple photochemical theory of visual acuity, which equates visual acuity to a form of intensity discrimination by the retina, is not entirely satisfactory. Examination of calculations, results, and assumptions, shows that it is both inadequate and in some details incorrect. The interpretations of the obtained data have been questioned, and alternative interpretations, which may better account for the data, have been offered.

Visual Acuity as Influenced by Neural and Central Processes

Marshall and Talbot's theories. One of the most recent attempts to account coherently for some of the rather contradictory data of visual acuity and of retinal anatomy and physiology is that of Marshall and

Talbot (35). For these authors, the retina is but the first step in a series of many; the impulses received by the brain are the ultimate determiners of visual perception, and many processes may intervene between the initial process of photochemical decomposition in individual receptors and the reception of impulses by the brain. They have attempted to explain the data of visual acuity in terms of these processes.

They state that sensory localization includes two categories of reactions: the minimal separation for two-point discrimination, which may be explained by the separation of end-organs with separate and discrete paths to the visual cortex; and second, contour discrimination and pattern recognition. The latter, they feel, cannot be explained by the reactions of individual units, but must be accounted for by the properties of *populations* of receptors and their associated neurons. They believe that the supposedly isolated nature of retinal elements and central connections has been overemphasized. They point to the evidence for neural summation. They do not emphasize the usual distinction between temporal and spatial summation, but speak instead of *lateral* and *vertical* summation as the two components of neural summation. *Lateral* summation refers to summation at the same neural level, as between bipolar cells, for example; *vertical* summation refers to summation between one neural level and a higher one, as between rods and bipolar cells. Lateral and vertical interaction processes work together, the lateral processes being of particular importance in the periphery of the retina and its associated neurons, and the vertical in the fovea. The lateral component integrates intensity from large numbers of receptors, and the vertical component aids in preserving pattern vision.

Altogether, these authors, in their brilliant but difficult article, suggest seven mechanisms which must be taken into account when explaining the various data of visual acuity. These are as follows:

1. Diffraction by the pupil (and, it might be added, scattering by dioptric media,) produces a statistical distribution of intensities from a point source of light.
2. Physiological nystagmus, which applies the graded distribution of intensities to separate receptors in a manner which is itself statistically distributed.
3. Reciprocal overlap between neural pathways, which is a mechanism for increasing the gradient of any excitation. It is further pointed out that the distribution of contacts is "peaked"; that is—in a submaximal reaction which involves only part of the available neurons, the greatest reaction density will be at the center of the reacting group. On a numerical basis, more contacts are involved there. Thus on account of this spatial distribution of contacts, and also on account of temporal summation, the increase of the gradient of excitation due to reciprocal overlap is especially stable.
4. The neural recovery cycle amplifies or depresses excitation, depending upon the part of the cycle during which the stimulus is applied. (Actually, Marshall and Talbot were not considering short exposure times, and it is possible that the times of the neural recovery cycle are also statistically distributed.)

5. Multiplication of the visual pathway. That is, each retinal receptor projects functionally, not to a single cortical cell, but to a probability distribution of cortical cells—not always by the same paths or to the same cells. Anatomical evidence indicates that in the cat and the *Rhesus* monkey at least, there is an increase in area and in volume as we ascend through the primary projection system. The relations of retina, geniculate, and cortex are conceived of as expanding cylinders, providing an increased volume of 1:10,000 from retina to cortex. Thus the cortical mosaic is far finer grained than is that of the retina.

6. Threshold mechanisms (possibly photochemical in nature) pass more or less of the pattern of activity.

7. A range of neural activity which covers about two log units of stimulus intensity. This range is thought of as operating at any given condition of adaptation, and as independent of thresholds. That is, at any given level of adaptation, a range of stimulus intensity covering two log units is thought of as mediated by neural mechanisms alone. Thus the total number of impulses arriving at the brain is thought of as a function of both the number of receptors active at that level, and the number of impulses delivered by each. The number of impulses delivered is determined not only by the stimulus intensity, but also by the way in which delivery is modified in the nervous system (*i.e.* facilitated, inhibited, peaked, and so forth).

Marshall and Talbot have obtained evidence for the existence of these factors from several sources; anatomical study, electrical investigation of the properties of the nervous system central to the retina, studies of eye movements, and behavioral data all indicate that such mechanisms are active. They consider various phenomena of visual acuity in relation to the interaction of the above factors, which might produce the obtained results. For example, with regard to the problem of how we can perceive a line whose width is much less than that of a single cone (although it must be considerably longer) several facts must be considered. As has been stated many times previously, the image on the retina of a hair-line is not the same as the simple geometrical image would be. Optical errors, such as diffraction by the pupil and chromatic and spherical aberration, diffuse the photic pattern. Light is scattered by the dioptric media. The net result is a distribution of light on the retina whose form has been calculated (19). Apparently the image is a normal distribution whose half-width is estimated to be 44" (for a 2.3 mm. pupil). This distribution is the combined result of blurring by the optical system and of optical diffuseness from physiological nystagmus.

Marshall and Talbot propose a "dynamic" interpretation of the relation between the size of the receptor element and the optical pattern. They point out that the receptor subtends the steepest slope of the distribution of light. Consequently, physiological nystagmus produces the maximum rate of change of light as the distribution of light intensity traverses the receptor. Smaller receptors, they say, would be useless, because, though traversing the optical gradient oftener, they would

gather proportionally less brightness differential. "The limiting retinal factor in acuity seems to be the relation of receptor width to the highest optical *gradient* in a moving pattern, rather than the average static differential illumination of one cone, compared with its neighbors" (35, p. 137).

A brightness difference of only 1% may be perceived, but there is no evidence that it would not be subliminal if it were not impressed suddenly (by the fast-moving image) on many receptor "rows." The suggestion is made that the fluctuating gaze sweeps the long edge over the receptors, whose subliminal effects add (20) to evoke a differential sensation.

Thus we have what Marshall and Talbot refer to as an "intermediate image," the properties of which are determined by optical blurring, physiological nystagmus, receptor size, and neural summation. This intermediate image is "projected by a 'neural lens,'" refocussing the retinal activity pattern onto the cortex, correcting and modifying its sharpness, contrast, and form. The perceived line is straight and sharp and it appears to be of high contrast. These effects are not explained at the retinal level.

Some sort of "averaging" process has apparently been going on in the projection system which has so modified the image that it bears a closer resemblance to the simple geometrical image of the stimulus object than to the modified retinal image. One factor which may operate to produce such an "averaging" is physiological nystagmus.

The slopes and amplitudes of the distribution may be modified still further by two other factors; first, the relation of the nystagmic movement of the image to the neural recovery cycle; and second, the relation of the intensity distribution to the recovery cycle. Near an edge, the light intensity falls off, and with it the frequency of impulses transmitted. Both factors cause neural amplification for impulses within the supernormal period, and subnormality for longer periods. For both reasons, propagated activity at an edge is peaked at the bright side and depressed at the dim side, thus enhancing gradient. Furthermore, the recovery cycle is related to neural peaking by a phase factor, for the different neurons near a point in a synaptic field will be in different stages of recovery when a burst of activity arrives. "This distribution of relative thresholds may be regarded as another statistical mechanism in the transmission, which insures continual availability of pathway for vision" (35, p. 141).

Thus these authors are able to suggest mechanisms which allow detail lost at the retina to be regained at the cortex. Particular attention is paid to the various types of visual acuity: hair-line, vernier, contour breaks, and two bright bars; and to the way in which the above-listed mechanisms may operate to account for the observed data in each case.

The complicated and involved systems of interactions which they

posit are of very great importance to those who are trying to explain some of the difficult and contradictory phenomena observed in making measurements of visual acuity. The theories offered by Marshall and Talbot, while probably not yet complete, are significant because they are among the first authors to present an analysis of present data which accounts for a large number of the known facts of visual processes rather than just a few. The role of diffraction by the pupil and aberrations of the eye has been discussed by many authors and is, in fact, given some consideration now by almost every author who discusses visual acuity; the importance of physiological nystagmus has been emphasized by Andersen and Weymouth (2), by Adler and Fliegelman (1), and by Weymouth, Andersen, and Averill (45); the possibility of summation or inhibition central to the retina has been referred to often by Crozier and discussed by Polyak (36); threshold mechanisms have been suggested as the basis for visual acuity, notably by Hecht and his collaborators. But so far none of these earlier authors has attempted to take into account all of these mechanisms in a theory which would explain the observed data of visual acuity. This, Marshall and Talbot have done.

Such a theory must be, by its very nature, complicated. If seven different factors are operative (and there are probably more than the seven suggested by Marshall and Talbot; for example, they have made no attempt to take wavelength into account) the number of possible interactions between them is very large indeed. Since the experimental results depend on this sort of interaction, and not on any simple predominance of one factor or another, we should expect to find apparent contradictions, resolvable only by a careful analysis taking into account the recognizable factors and their interactions.

The eye has too often been compared to a camera. The analogy is sufficient for some pedagogic purposes; but when we attempt to explain all visual phenomena in photographic terms we are leaving out two of the most important parts of the visual mechanism: the optic tract and the brain. A complete theory of any visual response must take into account the central as well as the peripheral components of the visual system.

SUMMARY

The chief problems which thus far have been investigated in the field of visual acuity concern the size of the minimum visible angle, and the effect of various stimulating conditions upon this minimum size. One of the most important of these variable stimulating conditions is intensity of illumination. Many of the values found for different types of targets have been considered, and it is pointed out that the order of magnitude of the measure depends to a very large extent upon the char-

acter of the visual target used. The main purpose of those who have made these measurements has been to determine the nature of the physiological processes underlying visual acuity. The old idea that the basis of resolution was the separation of two stimulated cones by one unstimulated cone has been shown to be inapplicable, and measurements a good deal finer than the width of a single cone have been presented.

Visual acuity varies in a systematic way with an increase in the intensity of the stimulating illumination, and this systematic variation must be accounted for by any complete theory of visual acuity. Two types of theories, with an illustration of each, have been discussed:

1. A theory of visual acuity based on peripheral processes. Hecht has been the primary recent proponent of this type of theory, but the concept of visual acuity as a form of brightness discrimination accomplished chiefly at a retinal level underlies most thinking on the subject today. (Cf. Bartley: (5, p. 34) "Visual acuity is a form of brightness discrimination in which spatial factors are the focus of investigation.") The picture most usually presented by the proponents of this type of theory is essentially a static one: the retinal image is conceived of as a (usually stationary) distribution of intensities. Some receptors are stimulated and others are not, because the brightness gradient at boundaries is sufficient to cause differential stimulation in some cases and not in others.

2. Theories which consider events in the retina, optic tract, and brain. The only authors who elaborate such a theory today are Marshall and Talbot (35). They have attempted to take into account all possible static and dynamic, retinal, and nervous processes which might influence the mechanism of visual acuity. Their theory is new, is by their own admission incomplete, and is not entirely documented by experimental evidence, but it represents a new type of thinking about an old problem and should prove fruitful for research.

In attempting to determine the basis on which an observer makes a discrimination of a small angular separation, it is no longer possible to consider nothing more than a distribution of intensities on the retina. Nervous processes, as discussed by Marshall and Talbot, are most certainly involved. It is the problem of further experimenters to investigate the nature and influences of these nervous processes by tests designed for that purpose.

BIBLIOGRAPHY

1. ADLER, F. H., & FLIEGELMAN, M. Influence of fixation on the visual acuity. *Arch. Ophth.*, 1934, 12, 475-483.
2. ANDERSEN, E. E., & WEYMOUTH, F. W. Visual perception and the retinal mosaic. *Am. J. Physiol.*, 1934, 18, 97-101.
3. AUBERT, H. *Physiologie der Netzhaut*. Breslau: 1865 (pp. 235-253).
4. AVERILL, H. L., & WEYMOUTH, F. W. Visual perception and the retinal mosaic. II. The influence of eye movements on the displacement threshold. *J. comp. Psychol.*, 1925, 5, 147-176.

5. BARTLEY, S. H. *Vision*. New York: Van Nostrand, 1941.
6. BAUMGÄRTNER, H. Die Formensinn und die Sehschärfe der Bienen. *Zeits. vergl. Physiol.*, 1928, 7, 56-143.
7. BERGER, C. The dependency of visual acuity on illumination and its relation to the size and function of retinal units. *Am. J. Psychol.*, 1941, 54, 336-352.
8. BERGER, C., MCFARLAND, R. A., HALPERIN, M. H., & NIVEN, J. I. The effect of anoxia on visual resolution. *Am. J. Psychol.*, 1943, 56, 395-407.
9. BEST, F. Kritische Bemerkungen zu Hechts Theorie der Sehschärfe. *Naturwiss.*, 1930, 18, 236-237.
10. BYRAM, G. M. The physical and photochemical basis of visual resolving power. I. The distribution of illumination in retinal images. *J. Opt. Soc. Amer.*, 1944, 34, 571-591.
11. CLEMMESSEN, V. Central and indirect vision of the light-adapted eye. *Acta Physiol. Skand.*, 1944, 9, Suppl. 27, 1-206.
12. CRAIK, K. J. W. The effect of adaptation on differential brightness discrimination. *J. Physiol.*, 1938, 92, 406-421.
13. CROZIER, W. J. The sensory discrimination of intensities. *Proc. Nat. Acad. Sci.*, 1936, 22, 412-416.
14. CROZIER, W. J., & WOLF, E. Theory and measurement of visual mechanisms. XI. On flicker with subdivided fields. *J. gen. Physiol.*, 1944, 27, 513-528.
15. CROZIER, W. J., WOLF, E., & ZER-RAHN-WOLF, G. Critical illumination and critical flicker frequency for response to flickered light in dragonfly larvae. *J. gen. Physiol.*, 1937, 20, 363-392.
16. DEL PORTILLO, J. Beziehungen zwischen den Öffnungswinkeln der Ommatidien, Krümmung und Gestalt der Insektaugen und ihrer funktionellen Aufgabe. *Zeits. vergl. Physiol.*, 1936, 23, 100-145.
17. FISHER, B. M. The relationship of the size of the surrounding field to visual acuity in the fovea. *J. exp. Psychol.*, 1938, 23, 215-238.
18. FREEMAN, E. What does a test of visual acuity measure? *Arch. Ophth.*, 1929, 2, 48-56.
19. FRY, G., & COBB, P. W. A new method for determining the blurredness of the retinal image. *Trans. Am. Acad. Ophthal. and Otolaryngol.*, 1935, 423-428.
20. GRAHAM, C. H. Some neural correlations. In C. Murchinson (Ed.), *Handbook of General Experimental Psychology*. Worcester: Clark Univ. Press, 1934. Pp. 829-879.
21. GRAHAM, C. H., & COOK, C. Visual acuity as a function of intensity and exposure-time. *Am. J. Psychol.*, 1937, 49, 654-691.
22. HARTLINE, H. K., MILNE, L. J., & WAGMAN, T. H. Fluctuation of response of single visual sense cells. *Fed. Proc. (Fed. Am. Soc. Exp. Biol.)* 1947, 6, 124.
23. HARTMANN, G. W. The increase of visual acuity in one eye through illumination of the other. *J. exp. Psychol.*, 1933, 16, 383-392.
24. HARTRIDGE, H. Visual acuity and the resolving power of the eye. *J. Physiol.*, 1922, 57, 52-67.
25. HECHT, S. The nature of the photoreceptor process. In C. Murchinson (Ed.), *Handbook of General Experimental Psychology*. Worcester: Clark Univ. Press, 1934. Pp. 704-828.
26. HECHT, S., & MINTZ, E. The visibility of single lines at various illuminations and the retinal basis of visual resolution. *J. gen. Physiol.*, 1939, 22, 593-612.
27. HECHT, S., PESKIN, J. C., & PATT, M. Intensity discrimination in the human eye. II. The relation between $\Delta I/I$ and intensity for different parts

- of the spectrum. *J. gen. Physiol.*, 1938, 22, 7-19.
28. HECHT, S., SHLAER, S., & PIRENNE, M. H. Energy, quanta, and vision. *J. gen. Physiol.*, 1942, 25, 819-840.
 29. HECHT, S., & WALD, G. The visual acuity and intensity discrimination of *Drosophila*. *J. gen. Physiol.*, 1934, 17, 517-547.
 30. HECHT, S., & WOLF, E. The visual acuity of the honey bee. *J. gen. Physiol.*, 1929, 7, 727-760.
 31. HELMHOLTZ, H. v. *Treatise on Physiological Optics*. Vol. II. (Translated from the 3rd German edition, J. P. C. Southall (Ed.).) Menasha, Wis.: George Banta Publishing Co. for the Optical Society of America, 1924.
 32. KOENIG, A. Die Abhängigkeit der Sehschärfe von der Beleuchtungsintensität. *Sitz. Akad. Wiss. Berl.*, 1897, 35, 559-575.
 33. KRAVCOV, S. V. Changes of visual acuity in one eye under the influence of the illumination of the other or of acoustic stimuli. *J. exp. Psychol.*, 1934, 17, 805-812.
 34. LYTCHGOE, R. J. The measurement of visual acuity. *Med. Res. Council, Spec. Rep. Ser.*, No. 173. London: His Majesty's Stationery Office, 1932.
 35. MARSHALL, W. H., & TALBOT, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. *Biol. Symp.*, H. Klüver (ed.), 1942, 7, 117-164.
 36. POLYAK, S. L. *The retina*. Chicago: Univ. Chicago Press, 1941.
 37. SHLAER, S. The relation between visual acuity and illumination. *J. gen. Physiol.*, 1937, 21, 165-188.
 38. SHLAER, S., SMITH, E. L., & CHASE, A. M. Visual acuity and illumination in different spectral regions. *J. gen. Physiol.*, 1942, 25, 553-569.
 39. STEINHARDT, J. Intensity discrimination in the human eye. I. The relation of $\Delta I/I$ to intensity. *J. gen. Physiol.*, 1936, 20, 185-209.
 40. TITCHENER, E. B. *A textbook of psychology*. New York: Macmillan, 1910.
 41. TONNER, F. Die Grösse der Empfindungsfläche eines Lichtpunktes und der Zapfenrastrer. *Pflüger's Arch.*, 1943, 247, 168-182.
 42. TROLAND, L. T. An analysis of the literature concerning the dependency of visual functions upon illumination intensity. *Trans. Ill. Eng. Soc.*, 1931, 26, 107-196.
 43. VOLKMANN, A. W. *Physiologische Untersuchungen im Gebiete der Optik*. Vol. I. Leipzig: Breitkopf and Härtel, 1863.
 44. WALLS, G. Factors in human visual resolution. *J. Opt. Soc. Amer.*, 1943, 33, 487-505.
 45. WEYMOUTH, F. W., ANDERSEN, E. E., & AVERILL, H. L. Retinal mean local sign; a new view of the relation of the retinal mosaic to visual perception. *Am. J. Physiol.*, 1923, 63, 410-411.
 46. WILCOX, W. An interpretation of the relation between visual acuity and light intensity. *J. gen. Psychol.*, 1936, 15, 405-435.
 47. WOLF, E. On the relation between measurements of intensity discrimination and of visual acuity in the honey bee. *J. gen. Physiol.*, 1933, 16, 773-786.
 48. WÜLFING, E. A. Über den kleinsten Gesichtswinkel. *Zeits. f. Biol.*, 1892, 29, 199-202.

AN ANALYSIS OF THE USE OF THE INTERRUPTION-TECHNIQUE IN EXPERIMENTAL STUDIES OF "REPRESSION"

ALFRED F. GLIXMAN

Department of Psychology, University of Mississippi

When three experiments are designed in much the same manner to answer the same question, and when they yield three different answers (5), then the time to examine carefully the techniques employed seems to be at hand.¹ The question asked in simple form is "Are there recall changes as a function of threat to self-esteem?" This question is basic to all experiments employing an interruption-technique in an attempt to test the psychoanalytic statement that one of the self-defensive reactions of man in moments of stress or frustration is to render the "painful" situation relatively inaccessible to recall (4; 3, p. 148f.). It is the purpose of this paper to examine both the rationale behind the use of the interruption-technique and the measures used to gauge the reactions produced.

The interruption-technique is a variation introduced by Rosenzweig (9) of the procedure employed by Zeigarnik (6, 7, 13). Zeigarnik, in her classical study of the recall of incompleting and completed tasks, prevented each of her subjects from completing half of the pencil-and-paper activities presented. After all the tasks had been presented, each subject was asked to recall as many of the tasks as possible. *The purpose of the experiment was to test the prediction that incompleting tasks would be recalled more frequently than would completed tasks.* Because this was the purpose of the experiment, the measure of the subject's response (recall) used was the ratio of the recall of incompleting to the recall of completed tasks. It is not the task of this paper to discuss the adequacy of such a measure or of Zeigarnik's statistical procedures in general; it is pertinent to note that the ratio used by Zeigarnik was perfectly in accord with the purpose of her experiment.

Rosenzweig's variation (7, 8, 9) of Zeigarnik's technique was based upon the induction of feelings of failure with respect to the incompleting tasks; i.e., all of the tasks were presented as a test of the subject's

¹ The experiments are those of Rosenzweig (9), Alper (1), and Glixman (5). Experiments by Rosenzweig and Mason (8) and by Sanford (10) will be omitted because children were used as subjects and because the raw data did not appear in the reports of the experiments. In general, the criticisms to be made of Rosenzweig and Alper also apply to Rosenzweig and Mason and to Sanford.

ability, and the instructions to stop working on any task indicated that S was performing poorly. *The purpose of the introduction of this variation was to produce an experimental setting in which one could study the effect of threat to self-esteem—i.e., the effect of "pain"—upon recall.* The difference between the recall of incompleting and completing activities was used as a measure of the response made by S to the experimental situation.² It should be noted that this measure is similar to that used by Zeigarnik in that it compares the recall of incompleting with that of completing tasks. Rosenzweig reasoned that if the recall of incompleting activities *relative to the recall of completing activities* in a stress situation—i.e., one that contains a threat to self-esteem—is less than the recall of incompleting activities relative to the recall of completing activities in a non-stress situation, then "repression"³ has been induced experimentally. The inadequacy of this measure is apparent immediately; any change in the recall of incompleting activities is a function not only of the increase in stress, but it is also a function of the change in the recall of the completing activities. A lowered recall-difference score (see footnote 6) may come about in a number of ways: (1) *decrease in recall of incompleting tasks*, with either no change or an increase in recall of completing tasks; (2) *no change in recall of incompleting tasks*, but an increase in recall of completing tasks; (3) *increase in recall of incompleting tasks*, with a greater increase in recall of completing tasks. The final test of the adequacy of a measure which is based on the comparison of the recall of incompleting with the recall of completing tasks lies in the effect such a measure has in the interpretation of the data. The remainder of this article is concerned with the analysis of the results of three experiments using an interruption-technique for the study of "repression." The interpretations of the results will be given, the summary results on which the interpretations are based will be presented, and the data will be re-analyzed in a manner which keeps the recall for incompleting tasks separate from the recall for completing tasks.

ROSENZWEIG'S EXPERIMENT

In "An Experimental Study of 'Repression' with Special Reference to Need-Persistent and Ego-Defensive Reactions to Frustration,"

² Although more than one kind of measure has been used, Rosenzweig and Mason, Rosenzweig, Sanford, and Alper all used a score which was a combination of the recall of completing and of incompleting tasks.

³ In this paper, "repression" (in quotation marks) will refer to decrements of recall induced by experimental procedures employing an interruption-technique, and repression (not in quotation marks) will refer to the clinically observed loss of recall as a defensive reaction to "painful" situations. The writer argues elsewhere (5) that the two are not synonymous.

Rosenzweig (9) compared the recall of thirty subjects in an informal (non-stress) situation—i.e., a situation which does not present a threat to self-esteem—with the recall of thirty subjects in a formal (stress) situation. As his conclusions are rather involved, they will be quoted at length:

It will readily be seen from Table 2 [Table 1, present paper] that subject for subject, memory in the informal group favored the unfinished tasks 19 to 7 . . . With the formal group, where pride had been aroused and ego-defense activated by failure, the memory results favored the finished or successful tasks 17 to 8.

. . . it is to be expected, on at least clinical and theoretical grounds . . . that there exist modes of ego-defense which differ from the sort of impunitive repression here considered in that they aggressively, i.e., intropunitively or extrapunitively, defend the ego by such mechanisms as displacement, isolation and projection. These mechanisms, however, would not necessarily involve the forgetting of unpleasant experiences; they might actually involve some rumination over them at the expense of the more successful experiences. If, in the present experiment, such additional defenses arose during the formal sessions, one could not expect the predominance of successes in recall—the alleged effect of repression as an ego-defense—to have been much more marked in the results than appears . . .

As regards the predominance of successes over failures in recall during the formal sessions, it is thus clear that this effect could have appeared only by overshadowing the countervailing effects of need-persistence and of certain aggressive types of ego-defense. The fact that such a predominance was actually found would seem to lend considerable support to the concept of repression as a very general mechanism of defense.

These considerations, however, tend to raise a doubt as to the significance of the present findings for the concept of repression itself. Theoretically repression should entail the conscious *forgetting* (and unconscious remembering) of unfinished, i.e., failed and unpleasant tasks; actually the results demonstrate that under the formal conditions, where ego-defense might be expected, the greater recall of successes is more striking than the forgetting of failures. This fact is doubtless attributable in part to the just mentioned countervailing effect of need-persistence related to the unfinished (failed) tasks; the shortcomings of the technique in identifying failure with incompleteness is again clear. In partial support of the present findings it can only be repeated that the psychologically sound approach requires the proportion of finished and unfinished tasks in the recall of each subject to be emphasized, not the absolute number of tasks of the two kinds recalled by the subjects collectively.

There is, moreover, a point of view from which the above noted defect of technique may instead be regarded as a virtue. As has been previously stated, repression theoretically involves not merely the defense of the ego from unpleasant memories but also the inhibition of certain interrupted activities. Such inhibited behavior may be conceived as persisting toward fulfillment, directly or indirectly, despite the existence of barriers. It is from such unsatisfied drives that, according to psychoanalytic theory, the conversion aspects of hysteria are energized at the same time that the ego defends itself by forgetting from the unpleasant images associated with the inhibiting trauma. If, now, in the present experiment, it is found that the forgetting of the unpleasant experiences is

complicated by a tendency for them to persist in memory because of their being incomplete, a closer approximation to the full concept of repression seems to be achieved than would be true if completed unpleasant tasks were in question.

The results upon which these conclusions are based are found in Table 1 (9, p. 69, Table 2). Table 1 contains the number of subjects in the informal and formal groups who recalled a preponderance of finished tasks, a preponderance of unfinished tasks, and an equal number of finished and unfinished tasks. A test for independence between kind of recall and experimental situation yielded $\chi^2=8.64$, $0.02>P>0.01$. It is clear, then, that a significantly greater number of people recalled a preponderance of completed tasks in the formal as compared with the informal group. In addition the mean ratio $100(\text{finished} - \text{unfinished tasks recalled})/(\text{finished} + \text{unfinished tasks recalled})$ for the informal group is -7.65 ; for the formal group it is 2.95 . "The indication thus is that the tendency for recalling unfinished tasks by subjects in the informal group was considerably more marked than was the corresponding tendency for those in the formal group to recall finished tasks" (9, p. 70).

TABLE 1
ROSENZWEIG'S (9) MEMORY RESULTS FOR GROUPS WITH CONDITIONS OF
NEED-PERSISTENCE AND EGO-DEFENSE

	Number of Subjects Who Recalled Preponderance of Finished Tasks	Number of Subjects Who Recalled Preponderance of Unfinished Tasks	Number of Subjects With No Preponderant Tendency
Group with informal conditions	7	19	4
Group with formal conditions	17	8	5

Rosenzweig's argument may be summarized as follows: The results of the experiment indicate that when the formal group is compared with the informal group there is a greater tendency to recall completed rather than incompleting tasks. Further examination of the results reveals that the most outstanding feature is the greater recall of the *completed* tasks in the formal situation. From Zeigarnik's work, a greater tendency to recall *incompleted* rather than completed tasks should be expected. There is, in addition, a slight decrease in the recall of incompleting tasks for the formal as compared with the informal group. The slight decrease plus the tendency expected from Zeigarnik's experiment suggest that repression is even stronger in Rosenzweig's experiment than is indicated by the results.

The point of view taken in the present paper is that both the ratio

described and the comparison of recall of completed with recall of incomplete tasks (upon which are based the value of χ^2) are inadequate. So long as recall of incomplete and completed tasks are compared with each other, there is no way of telling whether the change in ratios or the

TABLE 2

ROSENZWEIG'S (9) MEMORY RESULTS FOR INDIVIDUALS WITH CONDITIONS OF
NEED-PERSISTENCE AND EGO-DEFENSE

<i>Subjects with Informal (Need-Persistent) Conditions</i>					<i>Subjects with Formal (Ego-Defensive) Conditions</i>				
<i>Name</i>	<i>F*</i>	<i>U</i>	<i>F</i>	<i>U</i>	<i>Name</i>	<i>F</i>	<i>U</i>	<i>F</i>	<i>U</i>
	<i>Given</i>	<i>Given</i>	<i>Re- called</i>	<i>Re- called</i>		<i>Given</i>	<i>Given</i>	<i>Re- called</i>	<i>Re- called</i>
Bea	9	9	6	2	Ban	9	9	3	6
Ben	9	9	5	7	Bar	9	9	4	6
Ch	9	9	5	6	Br	9	9	7	6
Cl	9	9	4	6	Camo	9	9	3	2
Cra	9	9	7	7	Camp	9	9	8	8
Cre	9	9	3	5	Da	9	9	7	7
En	9	9	4	5	Dr	9	9	4	3
Fi	9	9	4	5	Fi	9	9	5	4
Fr	9	9	6	5	Gr	9	9	8	7
Gi	9	9	6	8	Ha	11	11	6	9
Ha	9	9	8	8	Ho	9	9	6	6
Hu	9	9	5	6	Ka	9	9	4	6
Ka	9	9	5	7	Le	9	9	4	5
Ke	9	9	9	4	Mc	9	9	5	4
La	9	9	6	6	Ni	9	9	6	5
Le	9	9	1	5	Pa	9	9	7	7
Ly	9	9	6	6	Por	9	9	5	3
Mc	9	9	4	7	Pow	9	9	7	8
Men	9	9	2	5	Smi	9	9	7	2
Mer	9	9	6	7	Smy	9	9	6	4
Mo	9	9	5	2	Sn	9	9	5	6
Ola	9	9	5	7	Ta	9	9	7	5
Olk	9	9	5	4	Trol	9	9	6	6
Po	5	5	5	3	Tros	9	9	5	6
Re	9	9	4	5	Wal	9	9	5	4
Sa	9	9	4	6	War	10	10	5	4
Un	9	9	7	6	We	9	9	5	4
We	9	9	3	4	Wh	9	9	8	7
Ya	9	9	3	6	Wi	9	9	5	9
Yo	9	9	5	7	Za	9	9	6	5
Totals	266	266	146	167		273	273	169	164

* F signifies finished, U unfinished tasks.

χ^2 cited is attributable to change in recall of incompleting tasks, recall of completed tasks, or both. Rosenzweig's data (Table 2, present paper; 9, p. 68, Table 1) have been analyzed for the change in recall of incompleting activities. The comparison of the mean recall in the informal situation with the one in the formal situation yields $t=0.07$; this is clearly not significant. The same comparison for the mean recalls of completed tasks shows a greater mean for the formal situation; $t=1.90$, $P=0.06$, for 58 degrees of freedom. It is apparent that no matter how Rosenzweig may choose to interpret his results, there is no decrement in recall of incompleting tasks when a stress situation is compared with a neutral one. There is, however, a nearly significant increment of recall of completed activities. The argument that the lack of selective forgetting is attributable to strong "need-persistent tendencies" is only to admit that the experimental design or the analysis of the data is inadequate, for this is to say that there is a hopeless confounding of variables. If the recall change for incompleting tasks is considered, Rosenzweig seems to have extracted two countervailing tendencies from thin air; on the basis of no change in recall of incompleting tasks, he has suggested that there are coexisting tendencies to recall and to forget incompleting activities. He indicates that the increased recall of completed tasks is an outstanding feature of the study, and then he dismisses this finding as a result of "the shortcomings of the technique"; in so doing, he has dismissed the one significant result of the study. If a decrement of recall is set up as a minimum criterion for repression, then there seems to be no justification for Rosenzweig's statement that he has achieved "a closer approximation to the full concept of repression . . . than would be true if completed unpleasant tasks were in question." In fact, if Rosenzweig's data are taken into account, he seems to have achieved no approximation at all to repression.

ALPER'S EXPERIMENT

The purpose of Alper's experiment (1) was to test the prediction that for "a given sample of subjects, unselected for personality factors, there will be no statistically significant differences between the incidental recall of completed and incompleting tasks if, experimentally, there is an equal number of completed and incompleting tasks to be recalled." The test was made by subjecting the same ten subjects to non-stress and stress conditions; therefore there is an implicit addition to the prediction to the effect that the lack of a significant difference between the recall of completed and incompleting tasks should hold for non-stress and stress situations.

Alper indicates that the "most important datum in Table 3 [Table 3, present paper] for present purposes is that *differences in selective recall*

TABLE 3

PERCENTAGE OF COMPLETED AND INCOMPLETED SENTENCES RECALLED BY TEN EXPERIMENTAL SUBJECTS UNDER THE NON-SELF-ESTEEM-INVOLVING CONDITIONS OF SESSION I AND THE SELF-ESTEEM-INVOLVING CONDITIONS OF SESSION II ADJUSTED FOR INDIVIDUAL DIFFERENCES IN PERFORMANCE (ALPER, 1)

Subjects	Percentage Recalled Session I		Percentage Recalled Session II	
	Completed	Incompleted	Completed	Incompleted
C	75	50	50	17
D	50	33	0	13
G	50	33	0	25
H	0	43	40	0
I	40	0	0	0
L	75	33	50	50
N	75	83	0	29
S	33	0	0	0
Ya	50	33	33	43
Yg	40	60	0	43

Difference between percentage completed and percentage incompleted recalled, adjusted for non-linearity of percentage score

	<i>t</i>	<i>P</i>
	(9 D.F.)	
Session I	1.52	.20-.10
Session II	0.65	.60-.50

within a given session were found not to be statistically significant." In discussing the importance of her results, she states (1, p. 414f.):

If we were forced to stay within the theoretical framework of Zeigarnik and Rosenzweig, we would have to say that the informal conditions of the first experimental session, Session I, tended primarily to arouse ego-defensive tensions (recall of completed tasks), while the supposedly more self-esteem-involving conditions of the second experimental hour, Session II, tended to arouse task-tensions (recall of incompleted tasks) in some S's, and ego-defensive tensions (recall of completed tasks) in others. Illogical as such an interpretation would be, on the basis of group data alone it can neither be refuted nor defended. It should be remarked, however, that the recall of incompleted tasks in a context of competitive failure need not be dynamically equivalent to the recall of incompleted tasks in a context of a "neutral" laboratory setting. In the same way, the recall of completed tasks under the two experimental conditions also need not be dynamically equivalent.

Alper then indicates that the discrepancy between her results and those of Zeigarnik (13) and Rosenzweig (9) might be attributed to individual differences in the response to the experimental instructions, but that analysis of group data alone "obscures these important individual differences here, just as it has in previous studies in this field." Alper takes the point of view that:

To focus one's successes when realistically threatened by failure may be the adjustive mechanism whereby immediate counteraction of failure is possible. To focus on one's successes in the absence of realistic failure threat may be a non-adjustive, non-integrative reaction symptomatic of low frustration-tolerance and inadequate counteractive mechanisms. The recall of incompleting tasks in an objectively unthreatening situation, as for example in Session I of this experiment, may be the "good" reaction of the secure, well-adjusted individual. The recall of incompleting tasks in an objectively threatening situation, as in Session II of this experiment, however, may be symptomatic of an over-readiness to admit defeat and of weak counteractive mechanisms.

Analysis which is to be published of the clinical data of the individual S's in this study justifies the interpretation given above. Correlation of independently obtained personality data with selective recall scores reveals that individuals who recall more incompleting than completing tasks in Session I, and more completing than incompleting tasks in Session II, can be characterized as Strong Egos. Individuals who recall more completing than incompleting tasks in Session I, and more incompleting than completing tasks in Session II, can be characterized as Weak Egos. The syndrome of personality parameters which characterizes the Strong Egos includes high Ego-Strength, high Conative Con-junctivity, high need for recognition, high need for dominance, low Dejection, Pessimism, and low Ego-Ideal, Intragression. Weak Egos rank high on Dejection, Pessimism, high on Ego-Ideal, Intragression, and low on Narcissism, on the need for recognition, the need for defendance, and the need for counteractive achievement. Moreover, when S's whose personality structure is such that they may properly be classified as Strong Egos, or as Weak Egos, are selected in advance of experimentation and then subjected to the two different experimental conditions of the present experiment, significant differences in selective recall are obtained in the expected directions. . . .

The results on which Alper's conclusions are based appear in Table 3 (1, p. 413, Table 3). The differences between the recall of completing and incompleting tasks are not significant for either the non-stress or stress situations; the *t*-value for the non-stress situation is 1.52 and the *t*-value for the stress situation is 0.65.

It should be noted that Alper's prediction explicitly involves a comparison between recall of completing and of incompleting tasks. The prediction states that for "a given sample of subjects, unselected for personality factors, there will be no statistically significant differences between the incidental recall of completing and incompleting tasks if, experimentally, there is an equal number of completing and incompleting tasks to be recalled." The term "differences" presents a major ambiguity. It either refers to the differences between the two kinds of recall within experimental situations, or it refers to the difference between the differences in the non-stress and stress situations. Each of these alternatives will be examined.

If major emphasis is placed upon recall differences within situations, then the use of two experimental situations seems gratuitous; in the absence of some other purpose, the use of either one of the situations

would have been sufficient. An additional purpose is indicated by Alper's promise to demonstrate that personality variables are related to patterns of recall. The experiment is then used as a test situation, and the "score" is a combination of recall reactions to both situations. The combination of differences between recall of incompleting and of completing tasks may be used to predict behavior of some other sort in some other situation. If it were Alper's purpose to use the two situations as a test situation, then no criticism could be made of the measure of reaction which is used. It should be noted, however, that Alper does not state this as the purpose of the experiment. In addition, it seems unlikely that a report of the results of a test situation of this kind would omit a statement of the predictive value of the resultant scores. The conclusion that the experiment was *not* designed as a test situation seems to be warranted; therefore, the use of the difference between the recall of incompleting and completing tasks within each of the situations is unjustified.

An alternative interpretation was given to the term "differences"; namely, "differences" refers to the difference between the differences in recall in the non-stress and stress situations. This interpretation implies that the major purpose of the experiment is to test the prediction that there is no difference between differences as a function of stress. Once the emphasis is shifted to recall changes as a function of stress, the inclusion of a stress situation becomes necessary, rather than a matter of choice. The justification for the shift in emphasis lies in Alper's conclusion:

Group data in support of Hypothesis I⁴ are presented. In contrast to the basic studies of Zeigarnik . . . and Rosenzweig . . . , in the present study no statistically significant differences in selective recall are found under either experimental instruction. It is argued here, not that selective recall is unlawful, but rather that the direction of recall is dynamically related to the self-esteem needs of the individual . . .

The reference to Rosenzweig and Zeigarnik in the conclusion inclines the writer to the point of view that the major purpose of the experiment is to study the changes in recall as a function of stress. This view is supported by the fact that a major portion of the introduction is allotted to a discussion of experimental approaches to repression and to the possible effects of "ego-involvement" on recall. A secondary purpose is to test the hypothesis that recall *changes* are a function of personality; therefore, if subjects are selected randomly with respect to personality variables, there should be no systematic changes in recall as a function of stress. It is unlikely that Alper is concerned with the

⁴ "Hypothesis I" is Alper's prediction which is stated earlier in this paper.

change in the difference between recall of completed and incompleting tasks when the stress situation is compared with the non-stress situation; the change is neither presented nor discussed (except to indicate that some subjects did change) in the analysis of the data. The fact that Alper presents Rosenzweig's experiments for comparison with her experiment further indicates that Alper was not interested in the difference between differences, for the differences between the recall of incompleting and completed tasks were of little concern to Rosenzweig. Rosenzweig used the differences in an attempt to demonstrate different recall changes for incompleting and for completed tasks as a function of stress. Alper's purpose, therefore, appears to be the same as Rosenzweig's, and her hypothesis may be stated in the following way: If recall changes for incompleting and completed tasks are largely a function of personality, and if a sample of subjects are selected randomly with respect to personality variables, then there should be no significant differences when either kind of recall in a stress situation is compared with the corresponding recall in a non-stress situation. The two kinds of recall are kept separate in the hypothesis to avoid Rosenzweig's position which implies a measure of recall that would obscure changes in both recall for completed and for incompleting tasks.

Again, one is faced with the question "Does the measure of recall make a difference in the conclusions drawn from the experiment?" Alper's conclusion was that no significant differences in recall had been found. If, however, the recall of completed activities in the stress situation is compared with the recall of completed activities in the non-stress situation, there is a significant decrease; $t = 2.84$, $P = 0.01$.⁵ If the recall of incompleting tasks in the stress situation is compared with the recall of incompleting tasks in the non-stress situation, there is a near-significant decrease; $t = 2.04$, $P = 0.07$. Thus, Alper's hypothesis must be rejected. There are systematic changes in the recall of completed tasks as a function of stress, even when subjects have been selected randomly with respect to personality factors.

GLIXMAN'S EXPERIMENT

The argument has been presented in this paper that the particular recall score used in the experiments on "repression" markedly affects the conclusions drawn from the experiments. In each of the two ex-

⁵ The policy of approximating as closely as possible the statistical procedures of other authors has been followed in the re-analysis of the data. Since Rosenzweig did not adjust his scores, no adjustment was made in the re-analysis. Alper used scores in the tests of significance which were adjusted: $\sqrt{X+0.5}$, where X refers to each of the percentages in Table 3. The t -test for matched groups was used. The same adjustment and test were used in re-analyzing Alper's data.

periments cited, a score based on the comparison of the recall of completed tasks with the recall of incompleting tasks had been employed; when the recall of completed tasks and the recall of incompleting tasks were treated *separately*, there emerged different conclusions from those offered by the experimenters. In order to present the issue at hand most clearly, a third experiment will be cited.

Elsewhere (5), the writer has presented an experiment designed to study the recall changes for incompleting and completed activities as a function of stress. For present purposes, it is sufficient to indicate that there were three situations representing three different degrees of stress; there were thirty individuals in each situation. Twenty paper-and-pencil activities were used. In Situation I, the activities were presented as tasks which were to be used in a later experiment; the subjects were being used to determine the length of time needed to complete the tasks. In Situation II, the activities were presented as an "Intellectual Alertness Inventory" which would be used to screen out potentially unsuccessful students; the subjects were part of a normative group. In Situation III, the activities were presented as the screening test which had already been standardized; the subjects' grades were to be correlated with the test performance. A group procedure was used to collect the data. Analysis of covariance (12) was used, with the effect of the number of completions partialled out. In effect, therefore, the scores employed were the number of completed and incompleting tasks recalled after an adjustment for the expected number (on the basis of the number of completions) had been made. Since the use of scores adjusted in this manner might obscure the comparison of the three experiments, the results based on scores more nearly like those used by the previous experimenters will be presented here. The following scores will be used: (1) recall ratio for

completed tasks $\left(\frac{\text{Completed Recalled}}{\text{Number Completed}} \right)$; (2) recall ratio for incom-

pleted tasks $\left(\frac{\text{Incompleted Recalled}}{\text{Number Incompleted}} \right)$; (3) recall-difference score

$$\left(\frac{\text{Completed Recalled}}{\text{Number Completed}} \right) \text{ minus } \left(\frac{\text{Incompleted Recalled}}{\text{Number Incompleted}} \right)^6$$

* Alper used the recall-difference score. Rosenzweig used $100(\text{finished} - \text{unfinished tasks recalled} / \text{finished} + \text{unfinished tasks recalled})$. Although Rosenzweig claims that this score is easily generalized to the case where the number of completions and incompleting are unequal for an individual, the resultant score is so unwieldy and its meaning so obscure that it was not used here.

Table 4 contains a summary of the relevant findings for recall ratios for completed tasks, recall ratios for incompleting tasks, and recall-difference scores, respectively. For each kind of score, the value of F for Situations is given.⁷ There is no significant variation among situation means for recall ratios of completed tasks: $F=2.44$; F at 5% point is 3.13. There is significant variability among situation means for recall ratios of incompleting tasks: $F=3.42$; F at 5% point is 3.13. On the basis of these results it is apparent that for the degrees of stress in this experiment recall of completed activities did not change as stress increased, and that recall of incompleting activities decreased as stress increased. If recall-difference had been used, the writer would have been forced to conclude that there was no change in recall as stress increased, for there was no significant variation among situation means for recall-difference scores: $F=0.50$; F at 5% point is 3.13.

TABLE 4

MEANS FOR RECALL RATIOS FOR COMPLETED AND INCOMPLETED TASKS AND FOR THE RECALL-DIFFERENCE SCORE OVER SITUATIONS I, II, AND III, AND THE F -VALUES ASSOCIATED WITH VARIATIONS AMONG THE MEANS*

	<i>Situation I</i>	<i>Situation II</i>	<i>Situation III</i>	<i>F</i>	<i>F at 5% Point</i>
Recall Ratio: Completed	0.60	0.62	0.52	2.44	3.13
Recall Ratio: Incompleted	0.57	0.52	0.44	3.42	3.13
Recall-Difference	0.03	0.10	0.08	0.50	3.13

* Glixman, from Tables II and IV filed as supplementary data to 5 with the American Documentation Institute.

Figure 1 indicates the trends of situation means for the three different scores. The mean recall ratios for completed tasks are: Situation I, 0.60; Situation II, 0.62; Situation III, 0.52. The mean recall ratios for incompleting tasks are: Situation I, 0.57; Situation II, 0.52; Situation III, 0.44. The mean recall-difference scores are: Situation I, 0.03; Situation II, 0.10; Situation III, 0.08. There is a general decrease for recall of incompleting tasks as stress increases, and as stress increases there is an increase followed by a decrease for recall of completed tasks. It is obvious, therefore, that the recall-difference score may obscure the

⁷ The data upon which these results are based are not given here. They may be derived from Tables II and IV filed as supplementary data to (5) with the American Documentation Institute.

trends of the component scores; in the experiment just cited, the recall-difference score did not reflect the significant change in recall ratios for incompleted tasks.

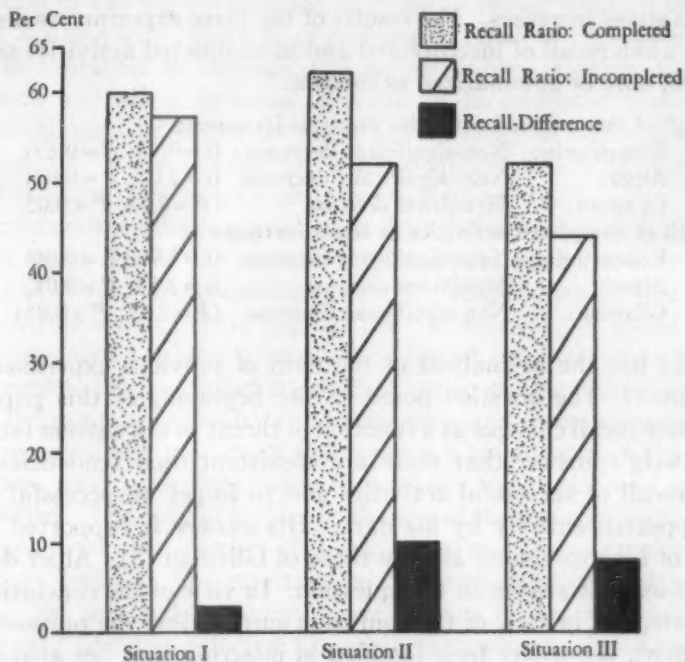


FIG. 1. MEANS FOR RECALL RATIO FOR COMPLETED AND INCOMPLETED TASKS AND FOR THE RECALL-DIFFERENCE SCORE: SITUATIONS I, II, AND III.

DISCUSSION

It should be emphasized that the issue presented in this paper is not statistical; it is based, rather, on the rationale underlying experimental studies of "repression." In an attempt to study the effects of stress on recall, using an interruption-technique, Rosenzweig (9) and Alper (1) have based their conclusions on scores which combine recall of incompleted and of completed tasks. Any statement about recall changes for incompleted tasks, however, implicitly involves recall changes for completed tasks. Evidence has been presented here which indicates that if recall of incompleted tasks is treated separately from recall of completed tasks, then conclusions other than the ones offered by the experimenters appear. Rosenzweig (9) suggests that an increase in recall of completed tasks *and* a decrease in recall of incompleted tasks result from an increase in stress. Re-analysis of the data reveals a near-

(significant change only for recall of completed tasks. Alper (1) implies that in a sample of subjects selected randomly for personality characteristics there are no significant recall effects as a function of stress. Re-analysis of the data indicates a significant decrease in recall of completed tasks as stress increases. The results of the three experiments discussed earlier, when recall of incompleted and of completed activities are kept separate, may be summarized as follows:

Recall of incompleted activities as stress increases:

Rosenzweig: Non-significant decrease ($t = 0.23$; $P = 0.82$)
Alper: Near-significant decrease ($t = 2.04$; $P = 0.07$)
Glixman: *Significant decrease* ($F = 3.94$; $P < 0.05$)

Recall of completed activities as stress increases:

Rosenzweig: Near-significant increase ($t = 1.90$; $P = 0.06$)
Alper: *Significant decrease* ($t = 2.84$; $P = 0.01$)
Glixman: Non-significant decrease ($F = 2.17$; $P > 0.05$)

What has the re-analysis of the data of previous experiments accomplished? The question posed at the beginning of this paper was "Are there recall changes as a function of threat to self-esteem (stress)?" Rosenzweig's answer that there are coexistent dual tendencies to increase recall of successful activities and to forget unsuccessful ones is not supported entirely by his data. His answer *is* supported by the results of his experiment and by those of Glixman (5). Alper does not state an explicit answer to this question. In view of the conclusions she does state, and in view of the confusion surrounding the purpose of the experiment, the writer feels justified in inferring that her answer is to the effect that there are no recall changes as a function of stress. Re-analysis of Alper's data indicates that the implied answer must be rejected. As a result of the re-analysis, previous answers have been modified, and the reasoning upon which the experiments were based has been made explicit. More important, sources of ambiguity have been removed from the interpretation of the data. Nonetheless, there still remains a surprising lack of correspondence among the results of the different experiments, even when the methods of analyzing the data are similar.

Suggestions have been made elsewhere (5) which would bring the results into agreement. Since the tasks in Alper's experiment permit a number of different solutions, and since the subjects were aware of this, the suggestion has been made that Alper's completed^{*} tasks really represented failures to her subjects. This suggestion could be tested by

* In Alper's experiment a completed task is one to which at least one solution has been found.

using different kinds of tasks under the same experimental conditions. If the suggestion is correct, then the results of Alper and the writer are in agreement. Rosenzweig's results were brought into agreement by the suggestion of the hypothesis that as stress increases through the lower part of a stress scale, there is an increase in recall of completed tasks; as stress continues to increase, this compensatory reaction disappears and there appears a decrease in recall of incompleted tasks. This hypothesis could be tested by using a wide range of stress points and breaking the "stress continuum"—a continuum which has still to be defined in other than very rough terms—into finer units.

A mistaken emphasis placed upon "the individual as an individual" appears to be one of the reasons for the errors committed by Rosenzweig and Alper. Both authors stress the importance of treating the individual as a unit. This approach led Rosenzweig to adopt a score which compared a subject's recall of completed tasks with his own recall of incompleted tasks. It led Alper to adoption of a similar score, to the use of the same individuals in the non-stress and stress situations, and to the proposal of a prediction which seems a bit unusual. Criticisms of the scores used have been made throughout the paper, and need not be repeated. The only objection to the use of the same subjects in both situations is that the number of experimental techniques employed may be unduly limited. The emphasis on "a given sample of subjects, unselected for personality factors" carries with it the implication that Rosenzweig had not selected his subjects randomly, for Alper never questions the significance of his results. The implication of decrying group results on the part of both experimenters is that the personality researcher must follow the supposed clinical practice of dealing with populations of one. As a matter of fact, it is doubtful that the clinician ever does this; if he did, his experience would be of no help to him. The outcome of the controversy between Sarbin (10) and Chein (2) seems to indicate that the clinician must deal with probability statements based, of course, on populations greater than one, but that these probability statements need not be made about resultant characteristics; they may, and often should, be made about "conditional" events.

Generally, it is recognized that a major function of personality researchers is to devise or discover situations in which it is possible to study the relationships among conditional events and between conditional events and resultant characteristics. The results of these studies must be some generalization which characterizes a group of events or subjects. Rosenzweig and Alper have stressed the importance of taking the individual's personality into account when their experiments are

evaluated. If the experiments were intended for the study of the relationship between recall changes and personality factors, the experimental designs should have included a personality classification of the subjects. The problem would then have been that of studying the recall behavior in the groups of subjects characterized by different personality patterns. Since there were no personality classifications, a safe assumption is that the purpose of the experiments was to study recall changes as a function of stress. For this purpose, stated in a simple form, the factors which make for individual differences are largely irrelevant; if significant changes take place as a function of stress, then one may generalize to a population which has the characteristics of the sample employed. Having found stress situations which yield recall changes, the experimenter may then relate other kinds of behavior to recall behavior. The writer feels that Rosenzweig and Alper were confused about the purposes of their experiments. Nonetheless, he feels that each of them has made an important contribution. Rosenzweig has provided an extremely fruitful technique for producing one kind of stress; Alper apparently has effected a thorough coordination of experimental attack and clinical investigation. It is unfortunate that these contributions should be clouded by a lack of clarity of purpose.

BIBLIOGRAPHY

1. ALPER, T. G. Memory for completed and incompleting tasks as a function of personality: an analysis of group data. *J. abnorm. soc. Psychol.*, 1946, 41, 403-421.
2. CHEIN, I. The logic of prediction: some observations on Dr. Sarbin's exposition. *Psychol. Rev.*, 1945, 52, 175-179.
3. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
4. FREUD, S. Regression, *Collected papers*, 4, 84-97. London: Hogarth Press, 1925.
5. GLIXMAN, ALFRED F. Recall of completed and incompleting activities under varying degrees of stress. (In press, *J. exp. Psychol.*)
6. LEWIN, K. Formalization and progress in psychology, *Studies in topological and vector psychology*, I. *University of Iowa studies, Studies in Child Welfare*, 1940, 16, 9-44.
7. PRENTICE, W. C. H. The interruption of tasks. *Psychol. Rev.*, 1944, 51, 329-340.
8. ROSENZWEIG, S., & MASON, G. A. An experimental study of memory in relation to the theory of repression. *Brit. J. Psychol.*, 1934, 24, 247-265.
9. ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistent and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, 32, 64-74.
10. SARBIN, T. R. The logic of prediction in psychology. *Psychol. Rev.*, 1944, 51, 210-228.
11. SANFORD, R. N. S. Age as a factor in the recall of interrupted tasks. *Psychol. Rev.*, 1946, 53, 234-240.
12. SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Collegiate Press, 1946.
13. ZEIGARNIK, B. Ueber das Behalten von erledigten und unerledigten Handlungen. *Psychol. Forsch.*, 1927, 9, 1-85.

THE TECHNIC OF HOMOGENEOUS TESTS COMPARED WITH SOME ASPECTS OF "SCALE ANALYSIS" AND FACTOR ANALYSIS¹

JANE LOEVINGER

Washington University School of Medicine

Psychological tests, whether measuring abilities, attitudes, or personality traits, commonly consist of a series of items each scored plus or minus. Plus may mean to do correctly, to agree with, to exhibit the characteristic, and so on. An important group of attitude tests is constructed so that for half the items agreement is scored plus and for the other half disagreement is scored plus. There are many ways in which item scores are combined to form total scores.

In contemporary psychometric practice, it is the rule rather than the exception that two people having the same score on a test will have scored plus on different items. When each score stands for a variety of patterns of pluses, it is difficult to conceive of these scores as measuring anything. Such scores are crude empirical devices known to have some predictive efficiency, but they cannot be called measurements in any strict sense.

If we wish to consider a set of test scores as measuring a psychological function or complex of functions, then all the items in the test must measure the function or functions. Cureton (5) has made the same point: "The most important requirement for a test whose scores are to be interpreted as measurements would seem to be that its items all draw upon the same set of abilities and traits." For such tests it will be the rule rather than the exception that two people with the same score will have the same pattern of pluses.

The aim of measuring one psychological function at a time has motivated a number of recent methodological researches in the field of test construction. The goal has been given a number of names, such as coherent or unified tests (16), uni-dimensional tests, univocal scores, scales, and homogeneous tests.

Part II of the present paper is concerned with comparing two approaches to uni-dimensionality, "scale analysis," developed by Guttman and others (8, 10, 11, 12), and the technic of homogeneous tests, developed by the writer (18). Guttman worked chiefly with attitude tests, but his results are stated in general terms. The technic of homogeneous tests was originally presented in terms of tests of ability. In

¹ I wish to thank Professors Arnold Rose and Raymond B. Cattell for critical reading of this paper in manuscript.

Part I of the present paper the idea of homogeneity is generalized to apply to psychological functions other than ability.

Isolation and measurement of the dimensions of personality and of ability have also been part of the aim of the older techniques of factor analysis. Some aspects of the competition and cooperation between factor analysis and the technic of homogeneous tests are explored in Part III.

I. THE LOGIC OF HOMOGENEOUS TESTS

The number of ways of combining item scores to form total scores is large and confusing. Logically, however, there appear to be just two major schemes, namely, counting the number of pluses and finding the median ordinal number of the items scored plus. When the median ordinal plus method of scoring is used, the method of ordering the items is an essential part of the scoring scheme. These two basic scoring methods have been elaborated with a variety of weightings, ratios, and other thumbnail devices. To simplify the discussion, we will begin by considering only items scored plus or minus and tests scored by counting the number of pluses or by finding the median ordinal plus.

Let us call tests all of whose items measure the same complex of functions homogeneous tests. For homogeneous tests, two people with the same score will have about the same pattern of pluses. There appear to be just two types of tests which will satisfy this requirement. The two types correspond to the two major ways of scoring, number of pluses and median ordinal plus. The first type will be called cumulative tests and the second type differential tests.

Cumulative Homogeneous Tests. In a perfectly homogeneous cumulative test when the items are arranged in order of decreasing popularity, each person from some defined population will score plus up to an item characterizing him and minus on all subsequent items.

In the case of tests of ability, clearly if two items measure the same ability, then the ability to do the harder presupposes the ability to do the easier item. When the items are arranged according to difficulty, everyone will succeed up to a certain item, the one characterizing his level of ability, and fail all subsequent items, provided we have succeeded in our aim of constructing a perfectly homogeneous test.

In order to generalize the idea of cumulative tests, substitute for difficulty the complementary concept, popularity. The popularity of the item will be the proportion of the group scoring plus on that item.

Obviously, the appropriate way to score such a test is to count the number of pluses. In a perfectly homogeneous test the score could equally well be defined as the highest ordinal plus. One cannot expect to make perfectly homogeneous tests. Apparently the effect of the residual heterogeneity is minimized by continuing to score according to number of pluses.

Differential Homogeneous Tests. In a perfectly homogeneous differential

test, there is an order of the items such that each person from some defined population will score minus on all items up to a point characterizing him, plus on succeeding items up to another point characterizing him, and minus on all subsequent items.

Although a good many attitude tests are composed primarily of cumulative-type items, the typical example of the differential type of homogeneous test is an attitude test. Think of the items of a test as a series of statements all relating to the same attitude. Call the two poles of the attitude "left" and "right." Each statement in this type of test can be thought of as characterizing a single position on the attitude continuum between extreme left and extreme right. Each person also has a characteristic position on this continuum. He will agree to those items, if any, which express his opinion exactly and also to statements which differ only slightly from his opinion, whether to the left or right. He will disagree with statements very far to the left or to the right of his opinion. Unless our rules for administering the test are such that each person is forced to agree to a fixed number of statements, people will differ in two respects, namely, their positions on the attitude continuum and their "thresholds of acceptance." Threshold of acceptance refers to how far an opinion can differ from one's own and still be agreed with. The threshold of acceptance cannot be measured in terms of the number of items agreed to unless there is substantial reason for believing the items are evenly spaced on the attitude continuum.

Regardless of various thresholds of acceptance, however, in a perfectly homogeneous test there will be an order of the items such that for each person there will be no gaps in the items agreed to. Each person will disagree with all items to the left of some point and all items to the right of some point and will agree with all items between those two points. No doubt many attitudes can be measured by either the cumulative or the differential type of test. The touchstone of a differential-type item is that those who disagree with the item may be either to the right or to the left of the position it represents.

The appropriate way to score differential tests is according to the median ordinal number of the items marked plus. Ordering the items is more important and more difficult than in the case of cumulative tests. Probably a method could be evolved for ordering the items purely on the basis of answers to the items, but the amount of work would be considerable. An alternative method is to make a trial ordering on the basis of common sense, relevant hypotheses, and available data. This ordering can then be improved by successive approximations, as follows: Score each person according to the first ordering. For each item obtain the median score of those scoring plus on that item. (There appears to be no weighty reason for preferring mean or median; so the median may be used for convenience.) The items can now be re-ordered, in accordance with increasing median score from the previous step. Re-score individuals and repeat. The process is automatically terminated when the order of the items no longer changes.

While the distinction between the two types of tests could be stated in even more abstract terms, most examples which come to mind are of a concrete type. Where such item patterns are found, they will usually reflect developmental sequences, normal or pathological. In general, developmental sequences can be expected to exhibit both differential

and cumulative aspects. For example, if we were measuring emotional maturity, we should note first that there are aspects of maturity which, once acquired, are seldom lost, and other aspects which represent phases of development later discarded. "Laces own shoe," "goes downtown alone," "uses family car" might be items in a cumulative test of emotional maturity. Preferred types of books, radio programs, and games might make up items in a differential test of maturity. Youngsters would reject items either too young or too old. The number and level of, say, radio programs listened to would depend on the child's level of maturity, his interest in radio programs as such, and the level of programs available. One who doesn't care much for radio may listen just to one or two programs which appeal exactly to his level, while the radio fiend will tolerate programs considerably too old or too young for his intellectual and emotional development.

Clearly the distinction between cumulative and differential tests is essential to the endeavor of constructing homogeneous tests of attitudes and other personality traits. If a cumulative test is ordered according to decreasing popularity of items and then scored according to median ordinal plus, the scores will be changed in magnitude but not greatly in order. On the other hand, very little can be expected by scoring a differential test by number of plus items. The worst and most likely possibility is that tests will be made of a mixture of items, so that no method of scoring is logically defensible. McNemar's (20) comprehensive review of methodology in the field of attitude testing mentions nothing corresponding to the distinction between cumulative and differential tests. An examination of currently used tests also confirms that many tests are composed of both types of items, though doubtless a similar distinction has been made in isolated contexts from time to time. One may reasonably expect a considerable improvement in many attitude tests and perhaps in personality tests, consequent upon noting this distinction, including only one type of item, and using the appropriate methods of scoring and ordering items.

II. "SCALE ANALYSIS" COMPARED WITH THE TECHNIC OF HOMOGENEOUS TESTS

Terminology

Guttman (10) has defined a "scale" as follows: "For a given population of objects, the multivariate frequency distribution of a universe of attributes will be called a *scale* if it is possible to derive from the distribution a quantitative variable with which to characterize the objects such that each attribute is a simple function of that quantitative variable."

This definition corresponds exactly to the criterion, two people with the same score must have the same pattern of pluses, and the criterion applies exactly to cumulative tests. In the case of perfectly homogeneous differential tests, two people can have the same score with slightly different patterns of pluses if they differ in their "thresholds of acceptance." Guttman does not distinguish the two types of tests, but the tests he has worked with are probably all of the cumulative type.

Guttman's use of the term "scale" has now the advantage of priority and quite wide acceptance, as compared to the term "homogeneous test." On the other hand, psychologists in the past have been divided between those who use the term scale for items combined on any principle or none, and those who use the term in the special connotation of "scaling." "Scaling" among psychologists refers to the process of substituting non-arbitrary or metric scores for the original scores, which are known to be partly a function of arbitrary editorial judgment. The problem of a metric is not solved by constructing Guttman-type scales, a fact which Guttman (10, and elsewhere) has acknowledged. The writer believes the term "homogeneous test" to be more suggestive of its precise meaning and more consistent with other psychological usage than the term "scale." Needless to say, the choice is a matter of individual judgment.

At some points the differences in terminology between scale analysis and homogeneous tests can be interpreted as reflecting differences in scientific philosophy. The term homogeneity is used only to describe the relationship between scores on two or more items. "Scalability," on the other hand, is conceived not as a property of a test as given, but as a property of a "universe of attributes," from which the test items constitute a sample. Guttman (10) says:

A basic concept of the theory of scales is that of the universe of attributes. . . . An attribute belongs to the universe by virtue of its content. The investigator indicates the content of interest by the title he chooses for the universe, and all attributes with that content belong in the universe. There will, of course, arise borderline cases in practice where it will be hard to decide whether or not an item belongs in the universe. The evaluation of the content thus far remains a matter that may be decided by consensus of judges or by some other means. This has been recognized before, although it need not be regarded as a "sin against the Holy Ghost of pure operationalism." It may well be that the formal analysis for scalability may help clarify uncertain areas of content. However, we have found it most useful at present to utilize informal experience and consensus to the fullest extent in defining the universe.

Guttman says further: "It is shown by the theory of scale analysis that *almost any* sample of about a dozen questions from the universe is

adequate to test the hypothesis that the universe is scalable, provided the range of content desired is covered by the questions" (11).

Strictly speaking, the hypothesis being tested, the scalability of the universe of attributes, is not a statistical hypothesis, as it is not formulated in terms of a probability law, and the method of testing is a series of thumbnail devices which have little in common with the rigorously deduced criteria properly called "tests of statistical hypotheses."

The terms "universe of attributes" and "universe of content" have no counterpart in the technic of homogeneous tests, and I am unable to find any meaning in them except what the investigator hopes to measure. Now quite possibly when we ask a soldier in six different ways whether he wants to stay in the Army, the investigator's judgment is practically infallible at determining whether all the items relate to the same general attitude. There may be many other such areas. But scale analysis is proposed as a perfectly general psychometric technic, and there are many areas of psychological testing where the investigator's judgment is far from infallible. Three possible errors the investigator can make are the composition of ambiguous items, inclusion of the wrong type of item, and inclusion of items which measure a different characteristic than he thinks. After facing dozens of irate students, I can testify that several semesters of intensive experience are not sufficient to insure against the composition of ambiguous achievement test items, those which will be interpreted in various ways by different students. One may mistakenly include a differential item in a cumulative test, or *vice versa*. A third source of error lies in the inclusion of items which measure perfectly well a trait other than the one the investigator intends. In this connection, one should remember that there is an important group of psychological tests that depends for its entire validity on the fact that even fairly shrewd and sophisticated subjects cannot guess the purpose of entire tests. How much easier to be mistaken about a single item! Even if we allow the consensus of experts as evidence for the relevance of an item, the investigator himself, however expert, is undoubtedly too biased for his opinion to be useful. To lay the incubus of infallibility on the investigator is, at very least, to discourage the discovery of new relationships among apparently dissimilar items.

The concept of homogeneity has been developed as an alternative to the concept of reliability, and the degree of homogeneity of a test, like the degree of reliability, is intended to be stated numerically. Guttman originally (10) classified universes of attributes as either scalable or non-scalable, but later broadened the dichotomy to a trichotomy, scalable, quasi-scalable, and non-scalable. This usage is comparable to

a proposal that all tests be classified as "reliable," "quasi-reliable," or "unreliable." There are at least two objections. What is reliable (or homogeneous, or scalable) enough for some purposes is not good enough for others. But more essentially, no scientific purpose can be served by introducing discontinuities into our vocabulary which do not correspond to discontinuities in our data. Guttman has not offered any evidence that the lines between scales, quasi-scales, and non-scales are drawn to correspond to gaps in the data.

Guttman's reply (13) to a similar criticism, that the distinction between a scale and a quasi-scale is a categorical distinction rather than a quantitative one, in no way improves his position. Lewin labelled this type of thinking as Aristotelian:

When the Galilean and post-Galilean physics disposed of the distinction between heavenly and earthly and thereby extended the field of natural law enormously, it was not due solely to the exclusion of value concepts, but also to a changed interpretation of classification. For Aristotelian physics the membership of an object in a given class was of critical importance, because for Aristotle the class defined the essence or essential nature of the object and thus determined its behavior in both positive and negative respects.

This classification often took the form of paired opposites, such as cold and warm, dry and moist, and compared with present-day classification had a rigid, absolute character. In modern quantitative physics dichotomous classifications have been entirely replaced by continuous gradations. Substantial concepts have been replaced by functional concepts (17, p. 4).

The outlook of a Bruno, a Kepler, or a Galileo is determined by the idea of a comprehensive, all-embracing unity of the physical world. The same law governs the courses of the stars, the falling of stones, and the flight of birds. This homogenization of the physical world with respect to the validity of law deprives the division of physical objects into rigid abstractly defined classes of the critical significance it had for Aristotelian physics, in which membership in a certain conceptual class was considered to determine the physical nature of an object.

Closely related to this is the loss in importance of logical dichotomies and conceptual antitheses. Their places are taken by more and more fluid transitions, by gradations which deprive the dichotomies of their antithetical character and represent in logical form a transition stage between the class concept and the series concept (17, p. 10).

The foregoing objections to the terminology of scale analysis bear summary restatement:

1. The term "scale" has irrelevant metric connotations to most psychologists.
2. The phrase "testing the hypothesis of scalability" has the connotation of tests of statistical hypotheses, but scale analysis methods are intuitive and not rigorous.

3. The "basic concept" of a universe of attributes means simply what the investigator hopes he is measuring, a definition which, indeed, sins against pure operationalism.

4. The terminology and methodology of scale analysis force the classification of tests and their corresponding "universes" into arbitrary classes regardless of the relative frequency of extreme and borderline cases; in contrast, the statistics of homogeneous tests will be seen to provide for quantitative distinctions and to impose no restrictions on the distribution of the homogeneity of tests.

Ordering of the Items

The order of the items, which is part of the definition of cumulative homogeneous and differential homogeneous tests, is not essential to the administration of the test, and in the case of cumulative tests, does not affect the scoring. It does affect the scoring of differential tests and the evaluation of the degree of homogeneity for both types.

Guttman (10, 13) has invented an ingenious mechanical device, called the "scalogram board," for accomplishing this ordering. He confesses that the board is somewhat expensive and cannot be used for more people or more items than are specified in its original construction.

In published illustrations of the use of the scalogram board, tests have been used which are of the cumulative type, with items scored plus or minus. A curious redundancy occurs in these illustrations, namely, that each item is represented twice, once giving those scoring plus, once giving those scoring minus. Exactly as much information is recorded if only those scoring plus are represented.

A more essential criticism is that in the case of cumulative tests, the optimal ordering is given directly by the popularity of the items. Ferguson (6) in his Table 2 drew up a checkerboard of item scores, representing people in order of increasing scores as the columns and items in order of increasing difficulty (decreasing popularity) as rows. On this checkerboard, all the pluses will lie above a broken, more or less diagonal line if the test is perfectly homogeneous. Blank spaces above this line and pluses below indicate deviations from perfect homogeneity, the farther away from the diagonal line, the greater the deviation. Guttman appears to have worked exclusively with cumulative tests, and for them, at least if items are dichotomously scored, everything that can be accomplished by the scalogram board is done more simply and more exactly by Ferguson's checkerboard.

The methods of scale analysis include two alternatives to the scalogram board, namely, the Goodenough technique (8) and the Cornell technique (11). Both techniques are designed for use with multiple choice items and need not be elaborated here. Guttman (13) states,

"For achievement tests, where all items are dichotomous—being marked either right or wrong—the Cornell technique is perhaps the best of all to be used." The Cornell technique is not, however, as straight-forward and efficient as the Ferguson checkerboard.

For differential-type tests, a person who has a scalogram board may very well find it an efficient device for ordering items. The numerical method for ordering items in a differential test proposed in the first part of this paper was evolved after reading Festinger's (7) description of the scalogram board. It is roughly a numerical equivalent of the mechanical process possible with the board. Certainly with a little practice the mechanical process must be quicker. The numerical method requires no special equipment and is not unduly laborious. The numerical method is superior in providing a clear criterion for optimal order. The scalogram board may thus be a useful piece of equipment, but it is of less value than originally claimed.

Possibly it is an injustice to the scalogram board to compare it with Ferguson's checkerboard for the case of dichotomously scored items, for Guttman worked mainly with multiple choice items. The logic of scale analysis, worked out for multiple choice items, applies immediately to dichotomous items; however, the logic of homogeneous tests, worked out for dichotomous items, requires some stretching for it to apply to multiple choice items.

For convenience, let us refer to each choice of a multiple choice item as a sub-item. There is no need for the number of sub-items to be constant for the items of a given test. Let us assume, as is most often the case, that each person scores plus on one sub-item in each item. It will now be true that each sub-item is dichotomously scored, but the scores of the sub-items within an item are related.

For differential tests, the sub-items may be treated as items without further ado. This type of differential test will differ from the originally described differential test in that everyone will score plus on the same number of sub-items, and there will be a mutually exclusive relation governing plus scores on sub-items within a given item.

For cumulative tests, the sub-items within a given item must be ordered from low to high. Those scoring plus on any sub-item will be those originally scoring plus on that sub-item and all those scoring plus on higher sub-items within the same item. The lowest sub-item in each item is discarded, as everyone is credited with a plus on it. Thus sub-items within an item automatically have the relationship of perfect homogeneity for the cumulative type of test.

These formalities for reducing multiple choice to dichotomous items are offered with the hope that they will enable persons working with actual data to compare scale analysis with the technic of homogeneous tests. The writer personally doubts whether multiple choice items have any advantage over dichotomous ones to offset the methodological difficulties in most contexts. One exception to this statement may be

taken where it is desired to utilize the multiple choice data for intensity analysis (12), a process for which there is no equivalent in the technic of homogeneous tests.

Descriptive Statistics

The decision as to whether a universe of content constitutes a scale, a quasi-scale, or a non-scale is made in part on the basis of a "coefficient of reproducibility." The logic of this coefficient is approximately as follows: If everyone with the same score has the same pattern of responses, then knowing the total score, one can reproduce all the responses of each individual. On the basis of a given set of data, the pattern of responses most closely corresponding to each total score is defined. Each score corresponds to a "scale pattern." The coefficient is then the percentage of all responses which fit the appropriate scale pattern, i.e., the percentage of all responses which are reproducible from the individual's score.

Festinger (7) has criticized this coefficient, as follows:

It is clear that applying a criterion like 85% or 90% reproducibility to all attempts at scaling, irrespective of the number of items involved or the number of possible answers to each item, leads to false conclusions. In one case where there are many items and many parts to each question, 85% reproducibility might be excellent consistency; in another case 85% reproducibility might represent no better than chance occurrence and be no evidence at all for unidimensionality.

In reply, Guttman (13) cited his previous writings to show that the coefficient of reproducibility was expected to be considered in relation to the frequency of response to each category of each item, the scattering of the non-scale responses, and the number of items. A rule of thumb for detecting spuriously high reproducibility is that where reproducibility is genuinely high, each category of each item should have more responses consistent with scale patterns than outside scale patterns. The coefficient of reproducibility has not been used as "the sole basis for drawing inferences from a sample of items. It is the basic one, because the reproducibility of the universe is essentially what is in question, but additional criteria have been and are being used."

Guttman's reply is not so much a defense of the coefficient as an admission that its deficiencies must be and are being taken into account. In drawing conclusions from the coefficient, the investigator is required to bear in mind several other factors, each of which consists of a number of *quantitative* observations. The coefficient of reproducibility is thus a highly inefficient statistic, based on only a small fraction of the relevant data.

In contrast to the coefficient of reproducibility, the coefficient of homogeneity takes into account all of the data, but it applies to a more limited group of tests. No doubt the restriction of consideration to cumulative tests composed of dichotomous items greatly simplified the problem of constructing efficient statistics to summarize the data. The defining characteristic of this type of test is that in the case of perfect homogeneity, the probability is unity of scoring plus on a more popular item for those known to have scored plus on a less popular item:

$$p_{i/j} = 1, \text{ for all } p_j \leq p_i,$$

where p_i is the probability of passing the i th item, and $p_{i/j}$ is the probability of passing the i th item among those known to have passed the j th item. The quantity p_i can also be interpreted as the popularity of the i th item. The coefficient of homogeneity is essentially a weighted average of the probabilities, $p_{i/j}$, for each pair of items, adjusted so that the coefficient will equal zero for a perfectly heterogeneous test and unity for a perfectly homogeneous test.

The worth of this coefficient depends not only on its logic but on the ease with which it can be computed, involving about as much work as the computation of two standard deviations. The most interesting step in the derivation of the computational form is the demonstration that the coefficient is a linear function of the variance of the test, with the constants of the function defined by the item popularities:

$$H_t = \frac{V_x - V_{het}}{V_{hom} - V_{het}},$$

where H_t is the homogeneity of the test, V_x is the test variance, V_{het} is the variance of a hypothetical perfectly heterogeneous test with the same distribution of item popularities, and V_{hom} is the variance of a hypothetical perfectly homogeneous test with the same distribution of item popularities. The quantities V_{het} and V_{hom} depend only on the item popularities.

Parenthetically, the assumptions underlying the coefficient have been incorrectly stated. The stated second assumption is the exclusion of negative homogeneity. While the coefficient is intended to discriminate degrees of positive relation, there is no necessity to exclude instances of negative relation. An assumption which is made but not stated is that, at least to a first approximation, the popularities of the items are independent of the context in which they are presented. This assumption will not always hold true; it will probably be closely approximated in well-constructed tests; and it can be tested by altering

the order of presentation of the items. In the case of tests formed from multiple choice tests, as described in the previous section, the assumption clearly does not hold, and the value of the coefficient for a perfectly heterogeneous test would be considerably greater than zero.

Festinger (7) objected to Guttman's distinction between scales and quasi-scales on the grounds that the distinction is arbitrary. It would be better, he states, "to content oneself with a description of the extent to which the scale on hand departs from the ideal of uni-dimensionality."

Guttman (13) replied that a universe which is quasi-scalable is just one of several kinds of non-scalable universes, and differs from a scalable universe in more than the degree of reproducibility. "The distinguishing feature is *the gradient in the responses to the items*. Cutting points cannot be established (as in the case of a scale) which will enable one to say that a person above the point is in one category of an item and a person below the point is in another category; but one can state that, if one person is higher than another in the quasi-scale, then his probability of being in a higher category of an item is correspondingly greater."

Here again, the technique of scale analysis requires of the investigator that he make a classification on the basis of a partially intuitive evaluation of a large amount of quantitative data. A quasi-scale is to be distinguished, for example, from a non-scalable universe which can be divided into two or more scalable universes. Guttman, on the basis of reports thus far published, apparently expects the investigator to bear in mind each person's answer to each item in making such a decision. Human intuition seems a weak instrument indeed to be entrusted with the decision as to whether a universe should be divided into sub-universes.

The technic of homogeneous tests provides two instruments for analysing the items of a test, a coefficient of the homogeneity of two items, H_{ij} , and a coefficient of the homogeneity of an item with a test, H_{it} . The coefficient of the homogeneity of two items can be defined as follows:

$$H_{ij} = \frac{p_{ijj} - p_i}{1 - p_i}, \quad \text{for } p_j \leq p_i,$$

where the symbols are defined as above.

This coefficient is algebraically equivalent to the one previously defined (18, p. 36). The original formula is a good working formula and shows that the coefficient is about as easy to compute as any well could be. From the above definition, obviously H_{ij} will equal zero for two

statistically independent items, in which case $p_i/j = p_i$, and will equal unity for two items in a perfectly homogeneous cumulative test, in which case $p_i/j = 1$.

The coefficient has two other interesting properties which were not mentioned in the monograph but can be verified with elementary algebra. It is equal to the ratio of the familiar "four-point r " or "phi coefficient" to its maximum value for the given item popularities. The coefficient of the homogeneity of a test, H_i , is a weighted average of the coefficients H_{ij} for each pair of items in the test. Combining these properties, we obtain the following relationship between the homogeneity of a test and the correlation between its items:

$$H_i = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^m p_j q_i r_{ij} / \max r_{ij}}{\sum_{i=1}^{m-1} \sum_{j=i+1}^m p_j q_i}, \quad \text{for } p_j \leq p_i.$$

The term r_{ij} indicates the four-point r between items i and j , and $\max r_{ij}$ indicates the maximum value of r_{ij} . As usual, q_i is equal to $1-p_i$. Other terms are as defined above.

The notion of dividing a coefficient, such as the four-point r , by its maximum value is so obvious that one suspects many people must have proposed it. Johnson (15), in fact, has done so, and also offered apologies to any unknown persons who may have anticipated him in the proposal. He showed that the ratio of r to its maximum value is algebraically identical to the ratios of his "coefficient of selectivity" and "coefficient of correctivity" to their respective maxima.

While the coefficients for measuring the homogeneity of a test and for measuring the homogeneity of two items are essentially similar, the coefficient of the homogeneity of an item with a test appears to be a different type of measure. Its logic is as follows: Every pair of individuals such that one scores plus on the item and one scores minus is discriminated by the item. Using as criterion the total score on the test minus that item, this discrimination is correct if the person scoring plus is higher on the test, wrong if the persons scoring plus is lower on the test, and not counted if the two are tied on the test. The percentage of correct discriminations minus the percentage of wrong discriminations equals H_{it} . Thus H_{it} equals one if all discriminations are correct, and equals zero if the correct discriminations exactly equal the wrong discriminations. This measure is an example of a general coefficient for measuring the relationship of non-metric variables (19), which is more

appropriate than Pearsonian correlation for a great many if not most uses in psychology.

For item analysis in the usual sense, namely, eliminating the least valuable items in a test, one would compute the m values of H_{it} , where m is the number of items in the original test. The amount of work is less than for computing the same number of biserial r 's or point biserial r 's. No rigorous connection has been established between H_i and H_{it} . The logic of the two coefficients is sufficiently similar so that one may reasonably assume that elimination of a few items for which H_{it} is markedly lower than for the others will raise H_i . If this statement cannot be supported algebraically, it will have to be verified empirically.

Some people have the impression from Guttman's writings that item analysis in the conventional sense is not permitted in scale analysis; once an investigator decides that an item is in the universe, there it stays. The following statement, taken from a context of derogatory remarks about item analysis, appears to support this view: "In scaling, we are interested in each and every attribute in the universe on its own merits" (10). In the same article, however, Guttman says, "It may well be that the formal analysis for scalability may help clarify uncertain areas of content." On the basis of Guttman's published writings, one cannot tell when scale analysis dictates the conclusion that a given set of items is drawn from a non-scalable universe, as opposed to the conclusion that certain items do not belong to a scalable universe from which the others are drawn. Guttman's specific criticisms of current methods of item analysis apply to Pearsonian correlation methods but not to the coefficient H_{it} .

If item analysis shows that the heterogeneity is distributed evenly over the m items, we still cannot decide whether we have what Guttman calls a quasi-scalable universe or two sub-universes. To make such a decision we need a table showing the $m(m-1)/2$ coefficients H_{ij} . Apparently for a quasi-scale these coefficients will all be moderate in magnitude and fairly uniform. In case the values of H_{ij} are some very high and some very low, despite uniform values of H_{it} , we expect there is a way of dividing the items into two or more tests each of which is more homogeneous than the original test.

Until a large experience is collected and recorded in terms of adequate statistics, there is no warrant for assuming that partially homogeneous tests will be readily sorted into piles corresponding to the above distinction, any more than that tests will sort themselves into piles labelled "homogeneous," "partially homogeneous," and "not homogeneous." Probably only rarely will we find clear-cut cases where no

selection of items will improve the test (all H_{ij} identical) or cases where the items separate easily into two or more highly homogeneous tests (all H_{ij} close to unity or to zero). The variance of the distribution of values of H_{ij} suggests itself immediately as a measure of where a given test falls between these extremes. Remembering that H_t is a weighted average of the H_{ij} 's, the variance of H_{ij} can be thought of as a measure of how far the homogeneity of the test can be improved by a different or by further selection of items.

III. FACTOR ANALYSIS AND HOMOGENEOUS TESTS

Are Homogeneous Tests Pure Factor Tests?

In the past half century traditional psychometrics, based on the work of Spearman, Pearson, and Binet, has contributed enormously to the development of psychology as science and as technology in the field of measuring abilities. In the field of measuring personality characteristics, there has been far from unanimous agreement as to the value of traditional psychometrics, and some of the most promising tests, notably the projective tests, have largely by-passed traditional procedures. Even with reference to measuring abilities, the pages of *Psychometrika* have contained an increasing number of articles in the past few years exploring the contradictions and dilemmas which arise from application of the concept of reliability, of the method of rectilinear regression, and of heterogeneity of test content (1, 2, 4, 5, 6, 9, 14, 23, 24, 25). Development of factor analysis has sharpened the appreciation of these difficulties.

The technic of homogeneous tests has been developed as an alternative to Spearman-Pearson-Binet psychometrics. "Sinking shafts at critical points" is replaced by the aim of homogeneity of content; the concept of homogeneity is shown to be a partial alternative to the concept of reliability; Pearsonian correlation is replaced by statistics appropriate to the data and to the aims of the tests.

The development of "pure factor" tests has long been an aim of factor analysis, within the framework of traditional psychometrics. The question naturally arises, what is the relation between homogeneous tests and the pure factor tests which factor analysis aims at producing? There are several answers.

Carroll (2) has stated as a property of a pure factor test of an ability exactly the criterion for a perfectly homogeneous cumulative test. He seems reluctant to claim, however, that the property is either a necessary or a sufficient condition for a pure factor test. He derived the same

formula as the writer did for the variance of a perfect homogeneous test. He also derived a formula showing that the magnitude of the Pearsonian correlation between two pure tests of the same ability could assume any positive value, depending on the distribution of item difficulties. An intuitive and informal proof of the same proposition is included in the writer's monograph.

Note, however, that nothing in the technic of homogeneous tests reflects the assumptions of one system of factor analysis as opposed to others. A test which is factorially pure in one system will not be so according to others. There is thus no formal reason for expecting homogeneous tests to be pure factor tests according to any factorial system.

The criterion for a cumulative homogeneous test will be satisfied equally well by tests composed of items which all measure a single factor and by tests composed of items which all measure an approximately constantly weighted sum of factors. Some people may find it psychologically more plausible that the investigator should succeed in constructing many items measuring a single ability, say, than in constructing many items measuring a constantly weighted composite of abilities. But factor analysis and not the technic of homogeneous tests is designed to distinguish objectively the factorially pure from the factorially composite test.

The interesting point is that the criterion of homogeneity more or less assures tests which measure an approximately constantly weighted sum of factors, and this is exactly the type of test which factor analysis assumes to begin with. The fundamental assumption of factor analysis states that for each test in the battery subject to factor analysis the score of each person is a weighted sum of his factor scores, with the weights constant for each test and with the factor scores the same for all people. Since different people do different items correctly, each of the items must depend on the same factors as the test as a whole and in about the same proportions.

Thurstone has discussed a closely related point: "Some writers have attributed to factor analysis the assumption that all the subjects in an experimental group use the same factors in doing a test, but such an assumption is not made in factor analysis" (22, p. 326). The discussion which follows this assertion, however, is not a discussion of the logic of factor analysis but an illustration of an instance where the assumption mentioned was not valid. In this instance, Thurstone says, "In order to make the analysis more complete, one might separate the subjects into two groups according to preferred methods of doing a test and reanalyze the results. In such a situation we should expect to find a different

factorial composition of a test for the two groups of subjects" (22, p. 326).

The two quotations from Thurstone are inconsistent. The equations of factor analysis do assume that the factor weights for a given test will be the same for all people. If people are to be separated into groups using different methods for the solution of a test, the separation is accomplished by some method other than factor analysis.

One may ask what will be accomplished by the technic of homogeneous tests in the situation described by Thurstone. Guttman (10) has suggested the use of scale analysis to pick out "non-scale types," that is, an occasional person who does not conform to the pattern of responses of the group. The separation of two more or less equal groups of people according to pattern of response is a different matter. It may perhaps be accomplished by development of an "inverted" technic of homogeneous tests, analogous to the Q-technique of factor analysis (21).

The practical value of a test ambiguous as to method of solution is questionable, however. Except in the improbable instance that the order of difficulty of items is independent of the factor or combination of factors used in their solution, the test will be revealed as not homogeneous and therefore unsuitable for factor analysis.

There is no reason to suppose that the factor analysis of tests helps directly in the composition of "pure" tests. Factor analysis of homogeneous tests will separate factorially pure from factorially composite tests according to the assumptions of the factorial system used. Can we not also say, the confidence, or perhaps the generality, to be attached to the results of any factorial study depends in large part on the degree of homogeneity of the original tests?

The Objective Definition of Psychological Traits

Factor analysis has been applied not only to tests but to items. In this application a closer comparison to the technic of composing homogeneous tests is possible. No exhaustive survey of the literature in this field will be attempted, but a few papers will suffice to show unsolved difficulties in applying factor analysis to items.

Ferguson (6) assumes that we start with a test composed of items homogeneous as to content but not as to difficulty. He describes these items as having the property which, in the language of the present paper, characterizes a cumulative homogeneous test. He admits that the correlation between two dichotomously scored items must be arbitrarily defined and does not attempt to justify his own choice of the

"four-point r " as the correlation between items. He then shows that factor analysis of the matrix of inter-item correlations will reveal not one factor, which one might expect on the basis of the homogeneity of content, but as many factors as there are levels of difficulty among the items. This finding suggests that some of the factors discovered in factor analysis not only of items but of tests may reflect not content differences but difficulty differences, and analysts may have gone astray in attempting to assign psychological meaning to such factors. He concludes that factor analysis had best be applied to batteries of items or of tests homogeneous as to difficulty.

Wherry and Gaylord (25) have answered that Ferguson simply chose the wrong coefficient as defining the correlation between items. The tetrachoric coefficient will be unity in the case of items in a test homogeneous as to content but not as to difficulty. The tetrachoric coefficient is thus the appropriate one to use for factor analysis of items; indeed, since the tests in most batteries subject to factor analysis vary as to difficulty, it would be best to dichotomize test scores and apply the tetrachoric coefficient here also. In a footnote, Wherry and Gaylord state that "one critic" objected that their use of the tetrachoric coefficient violates the assumption of normal distribution of scores on which the coefficient is based. "This is not a valid objection, however, since it is assumed that the trait (not the scores) is normally distributed. True, the trait *may not* be normally distributed."

Ferguson showed that if we use the four-point r as the correlation between items, the resulting factors may be due either to content or to difficulty differences. Wherry and Gaylord answered that using the tetrachoric coefficient eliminates the possibility of factors due to difficulty differences alone. Wherein has the use of the tetrachoric improved matters? For now we have no indication of how far the factors reflect the actual relationships between answers to items, and how far they reflect the assumptions underlying the tetrachoric coefficient, which are irrelevant and unverified in this context.

A further difficulty in the factor analysis of items is that as the number of items increases, the amount of work in carrying out the factor analysis becomes disproportionately great.

In contrast to factor analysis of items, let us consider what is involved in treating the same items by the technic of homogeneous tests. Assemble a large number of items, all dichotomously scored and judged to be of the cumulative type and all purporting to measure in the same general field, let us say, "personality." The number of items will be limited only by the patience of those taking the test. As no sampling

error formulas are available for the statistics of homogeneous tests, there should be not fewer than 100 subjects.

The items are arranged in order of difficulty and the coefficient H_{ij} computed for each pair. For this step punched card equipment and a calculating machine are helpful, but one can get along with no equipment but paper and pencil. A table of the values of H_{ij} is then drawn up. If the items are ordered according to difficulty, there will be entries only on one side of the principal diagonal.

The concluding step is to assemble the sets of items all of which have high inter-item homogeneities among themselves, for any set of items constitutes a test whose coefficient of homogeneity is a weighted average of the homogeneities of the items in pairs. Roughly, values of H_{ij} close to the principal diagonal are weighted more heavily than those farther from the principal diagonal. One may be able to constitute several homogeneous tests from a given original battery of items.

The foregoing procedure contrasts with factor analysis of items in the following ways:

1. The technic of homogeneous tests rests on fewer, more plausible, and testable assumptions.
2. For a given number of items, the method of homogeneous tests involves appreciably less work.
3. Psychologists can understand everything about the "theory" of homogeneous tests, if the methodology is worth the term "theory," with no more mathematics than high school algebra and statistical training not beyond the standard deviation.
4. The technic of homogeneous tests leads directly to the constitution of a test of predictable homogeneity. More often than not, factor analysis seems to lead not to tests but to hypotheses which it is hoped will lead later investigators to construct tests.

At this point the procedure of constructing homogeneous tests can also be contrasted with scale analysis. Guttman assumes as his first step in every case the definition of a "universe of content" and the selection of items within the domain so defined. According to the view taken here, which items go together can be determined entirely by the relationship between the answers to items chosen only for formal similarity, without any hypothesis on the part of the investigator. The further question of naming what is measured by a given homogeneous test is also not entirely a matter of the investigator's intuition. Relevant to the decision as to what is measured by a test are case studies of those receiving extreme scores on the test, and correlation of the given test with other measurements and ratings.

The procedure of starting with items chosen only for formal simi-

larity and without any hypothesis is not, however, recommended. Its importance is that it exists as a possibility. Much more reward for effort expended is to be expected if one starts with items chosen not only for formal similarity but because according to some scientific hypothesis they should conform to the criterion for a homogeneous test. In the field of personality, for example, radical theoretical differences exist as to what symptoms should be grouped together as a syndrome. Guttman's (13) statement that "neurotic phenomena have been found to be quasi-scalable" rather than scalable must be evaluated in terms both of the theoretical predilections of those drawing up the original tests and of the procedures, if any, used for item selection. In publications available to the writer, sufficient information is given on neither point.

In so far as a test possesses a high degree of homogeneity, it certainly measures something. In so far as we measure something, we are in a very real sense defining something. It need not be the case that all important psychological characteristics are so definable, nor that all characteristics so defined are of importance to psychology as science. The scientific achievement which a given test represents can be measured only partly by the coefficient of homogeneity; for a given degree of homogeneity, assuming a suitable distribution of item difficulties, clearly a test with greater absolute variance discriminates more degrees of the trait from each other. Guttman's (11) emphasis on the value of tests with a very few items may be a reflection of the *ad hoc* necessities under which scale analysis was developed. For general scientific purposes, tests with a considerable number of items seem as desirable as ever. Another factor in evaluating the scientific importance of a given test is the generality of the population for which the test holds up as homogeneous. In this connection Guttman (10) says, "Scales are relative to time and to populations," attaching no special significance to the generality of time or of population. Finally, other psychologists may agree with the writer that the apparent dissimilarity of the items is itself a criterion for the scientific importance of the test. When items are all phrased very similarly, the relationships may be purely a matter of semantics rather than of fundamental psychological characteristics.

Many of the terminological and methodological differences between Guttman and the writer are related to this difference: Guttman apparently has not conceived of scale analysis as a method of defining traits objectively; each investigator defines and names what he is measuring. The technic of homogeneous tests is proposed as a method for the objective definition of psychological characteristics. But many psycholo-

gists, notably Cattell (3), have accepted factor analysis as an adequate technique for the objective definition of traits. Are factor analysis and the technic of homogeneous tests in competition? Analysis of items by the technic of homogeneous tests seems to be alternative to treatment of items by factor analysis, but constitution of homogeneous tests and factor analysis of tests are complementary rather than competitive processes. In a few words one cannot do justice to Cattell's (3) extensive use of factor analysis to contribute to psychology as a science. Certainly his twelve "established primary traits" are more than just names, but they are less than measurable entities. The path from the isolation of primary factors to their measurement is not very clearly marked in Cattell's published research; however, in the case of many of his tests, which are direct measurements rather than item counts, the problems discussed in this paper do not arise.

To date, the elaboration of factor analysis as a group of methodologies has been entirely incommensurate with any accretion to psychology as a science resulting from studies employing factor analysis. One of the aims of factor analysis, the objective definition of psychological characteristics, appears to be more directly and somewhat differently served by the technic of homogeneous tests. The technic of homogeneous tests has the additional advantages of making fewer demands on the data and fewer demands on the investigator in the way of previous preparation. Whether an access in the objectivity of definition and measurement of psychological characteristics will result directly from application of the technic of homogeneous tests, or whether factor analysis will come into its proper scientific importance when sufficiently homogeneous tests are provided, remains to be seen.

SUMMARY

There are two types of psychological tests which satisfy the aim that all items shall measure the same function or functions. For *cumulative homogeneous tests*, when the items are ordered according to decreasing number of pluses, each person scores plus up to a characteristic item and minus on all subsequent items. For *differential homogeneous tests*, there is an order of the items such that each individual scores minus up to a characteristic item, plus up to another characteristic item, and minus on subsequent items.

The methods of scale analysis presented by Guttman have been compared with the technic of homogeneous tests. Both methods have been elaborated only for cumulative tests. The major points of difference between Guttman and the writer are as follows:

1. Scale analysis is primarily concerned with the "universe of attributes" underlying a test, while the technic of homogeneous tests is concerned in the first instance with the test.

2. Homogeneity is a quantitative attribute of a test, whereas in scale analysis the universe corresponding to a test is classified as scalable, non-scalable, quasi-scalable, or some more complex category.

3. The coefficient of reproducibility applies to a wider variety of tests than the coefficient of homogeneity. The coefficient of homogeneity has the value one for a perfectly homogeneous test and zero for a perfectly heterogeneous test, while the minimum value of the coefficient of reproducibility varies from test to test.

4. Methods of item-item and item-test correlation have been worked out appropriate to the construction of homogeneous tests, but articles on scale analysis are not clear on the admissibility of item analysis or on the methods to be used.

In discussing the relation of the technic of homogeneous tests to factor analysis, the following points were made:

1. Homogeneous tests need not be pure factor tests.
2. The equations of factor analysis assume that the tests in the initial battery are highly homogeneous.
3. Factor analysis of tests does not contribute in any simple way to the composition of pure tests of psychological functions. While factor analysis of items may do so, the technic of homogeneous tests has the advantages of avoiding unwarranted assumptions, of being less work, and of being conceptually simpler.
4. Factor analysis of tests and the technic of homogeneous tests can contribute to the objective definition of psychological characteristics in separate ways.

BIBLIOGRAPHY

1. BROGDEN, HUBERT E. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. *Psychometrika*, 1946, 11, 197-214.
2. CARROLL, JOHN B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, 10, 1-19.
3. CATTELL, RAYMOND B. *The description and measurement of personality*. New York: World Book Co., 1946.
4. CRONBACH, LEE J. Test "reliability": Its meaning and determination. *Psychometrika*, 1947, 12, 1-16.
5. CURETON, EDWARD E. Quantitative psychology as a rational science. *Psychometrika*, 1946, 11, 191-196.
6. FERGUSON, GEORGE A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, 6, 323-329.
7. FESTINGER, LEON. The treatment of qualitative data by "scale analysis." *Psychol. Bull.*, 1947, 44, 149-161.
8. GOODENOUGH, WARD H. A technique for scale analysis. *Educ. psychol. Msmt.*, 1944, 4, 179-190.
9. GULLIKSEN, HAROLD. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, 10, 79-91.

10. GUTTMAN, LOUIS. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 9, 139-150.
11. GUTTMAN, LOUIS. The Cornell technique for scale and intensity analysis. *Educ. psychol. Msmt.*, 1947, 7, 247-279.
12. GUTTMAN, LOUIS, & SUCHMAN, EDWARD A. Intensity and a zero point for attitude analysis. *Amer. sociol. Rev.*, 1947, 12, 57-67.
13. GUTTMAN, LOUIS. On Festinger's evaluation of scale analysis. *Psychol. Bull.*, 1947, 44, 451-465.
14. JACKSON, R. W. B. Note on the relationship between internal consistency and test-retest estimates of the reliability of a test. *Psychometrika*, 1942, 7, 157-164.
15. JOHNSON, H. M. Maximal selectivity, correctivity and correlation obtainable in 2×2 contingency-tables. *Amer. J. Psychol.*, 1945, 58, 65-68.
16. KELLEY, TRUMAN L. The reliability coefficient. *Psychometrika*, 1942, 7, 75-83.
17. LEWIN, KURT. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.
18. LOEVINGER, JANE. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, 61, No. 4.
19. LOEVINGER, JANE. The correlation of non-metric variables. Unpublished MS.
20. McNEMAR, QUINN. Opinion-attitude methodology. *Psychol. Bull.*, 1946, 43, 289-374.
21. STEPHENSON, WILLIAM. The foundations of psychometry: Four factor systems. *Psychometrika*, 1936, 1, 195-210.
22. THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univ. Chicago Press, 1947.
23. TUCKER, LEDYARD R. Maximum validity of a test with equivalent items. *Psychometrika*, 1946, 11, 1-13.
24. WHERRY, ROBERT J., & GAYLORD, RICHARD H. The concept of test and item reliability in relation to factor pattern. *Psychometrika*, 1943, 8, 247-264.
25. WHERRY, ROBERT J., & GAYLORD, RICHARD H. Factor pattern of test items and tests as a function of the correlation coefficient: content, difficulty and constant error factors. *Psychometrika*, 1944, 9, 237-244.

SUBJECT AND OBJECT SAMPLING—A NOTE

KENNETH R. HAMMOND

University of Colorado

There recently appeared two articles in this JOURNAL which dealt with sampling in psychological research (3, 4). The articles criticized a certain lack of sophistication in current procedures and outlined more mathematically refined methods. This is in line with a general trend which is laudable enough. There are, however, certain broader theoretical points to which we might give equal thought while proceeding toward greater mathematical precision. I refer to Brunswik's remarks in which he urges representativeness of *both* "situation" and population in the design of psychological experiments (1, 2). The procedure here will be to state briefly the essence of his remarks and then to criticize an experiment with respect to this position, thereby demonstrating its practical implications.

Brunswik's main point in connection with sampling is that it should be performed with respect to both subject and object, if we are to generalize in both directions. For example, in the ordinary social perception experiment a group of subjects, usually wishfully assumed to be a sample of some population, judges a social object, a person, say, for such a trait as likeability. Now ordinarily, generalizations from such a study depend largely upon the adequacy of the sample of judges—the more adequate the sample, the more faith in the generalizations. This is termed the "populational generality" of the result by Brunswik. But, he points out, we may well ask what sort of results we may expect when we ask our population to judge a different person, or object. In short, what of the "situational generality" of the results? Psychologists are eyeing sampling procedures more critically—but the criticism remains one-sided. The attempt is made to become more and more precise concerning the representativeness of samples of the populations to which we submit our objects to be judged, or tests to be taken, etc. But how representative are the objects, or the tests? Logic would demand equal representativeness on both sides of the experiment, for psychologists generalize constantly not only from populations of subjects, but from situations. What, otherwise, is the purpose of a testing procedure if not to generalize from the test situation?¹

¹ See R. C. Tryon (6) for a complete expression of the concept of sampling in connection with tests. For example, "The method (of testing) is that of *sampling* behavior, and it definitely presupposes that for any defined domain there exists a *universe* of causes, or factors, or components determining individual differences" (p. 433).

To sum up, psychologists maintain a one-sided emphasis on the need for representativeness. They emphasize representativeness of populations, but not situations, tests, or objects—thereby implicitly limiting the generalization of results obtained to the population, or subject, side.

In an effort to illustrate the consequences of onesidedness, we cite the experiment of Robinson and Rohde (5), in which the authors have "Jewish-appearing" and "non-Jewish-appearing" interviewers poll a population on a question bearing on anti-Semitism, in an attempt to determine the effect of the appearance of the interviewer on the responses.² As is customary, they go to considerable length to obtain a representative sample of interviewees (subjects), being careful to match the proportions in the sample to those in the population, use a large sample (2000), and compare responses of different strata to the "Jewish-appearing" and "non-Jewish-appearing" interviewers. They obtained significant differences in responses. But responses to what? Exactly two sentences are devoted to the description of the procedure of selecting the "Jewish-appearing" and "non-Jewish-appearing" interviewers, and exactly *nothing* is told about the sample of these interviewers (objects) which are presented to the population, i.e., how many, what sex, age, socio-economic status, etc. It is exactly as if we reported the responses of a group of subjects, but failed to report the stimuli to which they are responding, other than that they were persons (number unspecified!) *guessed* to have certain characteristics. As for the selection process, the authors report "A brief discussion was held with the whole group of interviewers on the matter of facial stereotypes of "Jewishness." Then by a majority vote, the interviewers were placed into either a Jewish-looking group because they fitted some of the stereotypes of "Jewish" appearance, or a non-Jewish-looking group because they did not fit in with these stereotypes." If this was considered good scientific procedure for object representativeness, or sampling, why was it not good enough for the subject, or population sampling? Precisely because the experimenters wish to generalize to the population—in this case, of New York City. But how far can they generalize with respect to "Jewish-appearing" and "non-Jewish-appearing" interviewers? No further than their sample of interviewers allows them. Since we know nothing of the sample, but from the selection process know that it is by no means a random sample of anything, we can therefore say only that the responses of the sample of the New York City population to

² It should be emphasized that this is not meant to be a criticism of this experiment in particular. There are many others that could have been chosen; this one merely happens to exemplify extremely well the point under consideration.

these interviewers are estimated to be such and such. In all likelihood New York City will never meet such a group of interviewers again. We therefore cannot generalize as to how this population will answer when presented with another group of interviewers, "Jewish-appearing" or "non-Jewish-appearing," no matter how they are selected. In short, the conclusion that "respondents express anti-Semitic opinions more readily with non-Jewish-appearing than with Jewish-appearing interviewers" is invalidated through failure to establish a representative sample of the situations (objects) to which the authors generalize. The nature of the independent variable is obscure.

If at this point the reader considers the criticism a minor one, the writer has failed to put across his point. The issue in question is the *validity of generalization*, hardly to be considered minor. We are criticizing the experimenters' generalization with respect to their interviewers, or objects. They have generalized to a population which they have by no means sampled. What could be more illustrative of our point that this is a common methodological error than the fact that they do not even discuss, let alone attempt, such sampling?

To pursue the point further, one should mention that from Brunswikian theory the sampling in this case was not only one-sided but it was on the *wrong* side. According to Brunswik (2, p. 37) "... proper sampling of situations and problems may in the end be more important than proper sampling of subjects, considering the fact that individuals are probably on the whole much more alike than are situations among one another."

To return to the experimenters' primary problem which was to discover whether or not the interviewers' appearance would influence interviewees' responses—the crux of the matter is to make certain the establishment of an independent variable. In this case it seems clear that the subjects can be more or less representative of this (hypothetical) population of "Jewish-appearing" or "non-Jewish-appearing" people. Sampling of this population (and of "situations," or poll questions) is necessary since there is in all likelihood an infinite series of combinations of appearances and situations, or questions.³

We have already belabored the point that nothing whatsoever is known here concerning the interviewers (objects) who represented this population. The population of interviewees (subjects) however, was apparently expertly sampled in order to generalize. But why generalize on *this* side of the experiment? It is convenient to have an estimate

³ We will not consider here the issue of question sampling. It is thoroughly covered in Tryon's article (6).

of how the New York population would have responded to these interviewers but it is far less relevant for this experiment than knowing to what they were responding. On the other hand, after having established an independent variable, had the authors exposed it to a small sample (perhaps 1/20 as large as they used) of an important segment of the population, it would seem a reasonable assumption that in the event differences appeared, one would be justified in pointing out the hazard of using interviewers whose appearance is correlated with the content of the question. The variable in question is thereby demonstrated to have been important to that sample, at least. To what other segments of the population it would be important is additional, worthwhile information, but far less important and not crucial, or particularly relevant, to this experiment.

Briefly, then, if one were to choose on which side of the experiment representativeness was most desirable, it seems clear that representativeness is mandatory on the stimulus, or object side, for that is where the experimenters *must* generalize in the problem in question—less necessary on the response, or subject, side. Not only, therefore, does this experiment demonstrate lack of representativeness in connection with the independent variable, but also a complete misplacement of effort in the direction in which representativeness was obtained.

To summarize, the purpose of this note was not to criticize the experiment cited, but to illustrate the practical and theoretical disadvantages which stem from what Brunswik calls the "traditional double standard" of representativeness in psychological experiments.

BIBLIOGRAPHY

1. BRUNSWIK, E. Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, No. 254.
2. BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Univ. of Calif. Press, 1947.
3. MARKS, E. S. Selective sampling in psychological research. *Psychol. Bull.*, 1947, **44**, 267-275.
4. MARKS, E. S. Sampling in the revision of the Stanford-Binet scale. *Psychol. Bull.*, 1947, **44**, 413-434.
5. ROBINSON, D., & ROHDE, S. Two experiments with an anti-Semitism poll. *J. abnorm. soc. Psychol.*, 1946, **41**, 136-144.
6. TRYON, R. C. A theory of psychological components—an alternative to "mathematical factors." *Psychol. Rev.*, 1935, **42**, 425-454.

ABSOLUTE PITCH—A REPLY TO BACHEM

D. M. NEU

Pennsylvania State College

In defense of an objective approach to the acquisition of behavior, several comments must be made about Bachem's (1) rebuttal of my article (2) on absolute pitch. Bachem rejects the view that absolute pitch is acquired, and is therefore the finest degree of pitch discrimination, to defend the view that absolute pitch is some entity inherent in the individual's behavior. He rejects the idea that absolute pitch can be learned and lists a number of statements which he says support his view.

Bachem states: "Pitch is the psychological counterpart of the frequency of air vibration." Taking the implications of this statement broadly, it is this psychological counterpart that we must consider in order to understand the development of absolute pitch. Tone height and chroma, as defined by Bachem, are physical components and are not relevant factors in our study.

Secondly, Bachem's psychological definition is: "Absolute pitch is the ability to recognize (and identify) the pitch of a tone without the aid of a reference tone." What is the difference between his definition and mine: "... the ability to discriminate tones without the aid of other tones to such a degree that naming or pointing out the note is rarely incorrect"? There is no difference here and nothing in either one that indicates a necessity for some inherent quality.

Bachem's third statement says: "Absolute pitch cannot be treated as an entity since absolute pitch identification is possible by different methods." These different methods result in different "types" of absolute pitch which he calls pseudo-absolute, quasi-absolute, and genuine absolute. The first two types are learned or can be improved "by experience and intentional learning," and the third type, genuine absolute pitch, is not learned and is therefore inherited, according to Bachem. I have found no evidence for saying that two types of pitch are learned and the third is a gift. We could give names to as many different degrees of pitch discrimination as we might choose, but this is not the problem. The point is that the studies in the literature do not produce any reasons for arbitrarily saying that some are acquired and some are inherited. Why not say they are all merely degrees of learning to discriminate absolute pitch?

Then, too, Bachem believes that an individual's having absolute

pitch, without ever attempting to acquire it, is evidence that the ability is innate. This belief does not agree with experimental facts. Learning for the most part is not a deliberate thing; it is going on constantly, and it is aided and abetted by all sorts of factors such as past experience, health, surrounding conditions, and changes in the relationships between the stimulus and the individual. With absolute pitch families we must consider many factors like similar surrounding conditions, imitation of behavior, similar opportunities to contact stimuli, and so forth. Learning and realization can be two different things. A person can build up very fine degrees of discrimination and only suddenly realize it when it is called to his attention by some situation or some individual. Certainly, as Bachem agrees, absolute pitch can be as "spontaneous as the recognition of colors," but experimental evidence does not show that colors are not learned either.

One more point must be considered. Failure to train an individual to have absolute pitch does not mean that it is inherited. If we take the view that it is inherited, we are completely overlooking all the evidence showing that the behavioral development during the earliest years of an individual's life is the most important and is more strongly ingrained. In the case of discrimination, if the individual develops a crude, less accurate, sort of way of attending to and discriminating between stimuli, he may never be able to overcome it in his later life.

Although there is no conclusive evidence to prove or disprove the inheritance of absolute pitch or any other kind of discrimination, I maintain that experimental results do not disagree with the above points. Since such a conception of absolute pitch represents an objective interpretation, I see no reason why it should not be considered as an hypothesis. In this way, we can approach the phenomena operationally and confine ourselves to a consideration of the experimental design that will constitute a crucial test of the hypothesis.

BIBLIOGRAPHY

1. BACHEM, A. Note on Neu's review of the literature on absolute pitch. *Psychol. Bull.*, 1948, 45, 161-162.
2. NEU, D. M. A critical review of the literature on "absolute pitch." *Psychol. Bull.*, 1947, 44, 249-266.

REPLY TO POSTMAN¹

G. RAYMOND STONE

University of Oklahoma

In my original note (3) on Postman's review of the literature on the law of effect (1) I juxtaposed what I thought to be representative samplings of quotations from Postman and Thorndike with reference to Thorndike's position on the influence of punishment. Without exception, it seemed to me, Postman had done Thorndike an injustice. In the second sentence following the quoted selections I wrote:

That the former (i.e., Postman) quotes the latter (i.e., Thorndike) correctly as to the indirect action of punishment in the elimination of responses (p. 502) is additional reason for surprise when he cites the long list of papers most of which, if not all, are irrelevant or at least not crucial to Thorndike's position (3, p. 153).²

It is this correct quotation (1, p. 502) that Postman requotes in his recent discussion (2) and claims for it that it dissipates the apparent contradiction between himself and Thorndike and answers virtually all of the points of my note. For my part I consider that it does neither of these things, but, instead, emphasizes the original reasons for the note.

I have little interest in perpetuating what Postman calls a "battle of quotations" and the present reply would have been deemed unnecessary had not Postman gone on to impute to me a statement which on the face of it is absurd.

In his original review Postman cited a long list of articles covering a half century of experimental work which he interpreted as being opposed to Thorndike's view of punishment. The implication that Thorndike had overlooked fifty years experimental work on punishment can only be considered ingenuous. As a matter of fact, in my note I had pointed out how Thorndike had analyzed some of these studies and related them to his own conclusion, implying, reasonably, I think, that all such studies could also be so related.

These studies were, therefore, irrelevant or not crucial to Thorndike's view of punishment and could not be used to refute him as Postman attempted to do. This is the meaning of my quotation at the beginning of this paper, but Postman handles this discussion in the following manner:

¹ Postman, Leo. Discussion of Stone's note on the law of effect. *THIS JOURNAL*, 1948, 45, 344-345.

² Italics added at the present writing.

My concern was with fitting Thorndike's view into the general picture of current learning theory and I must disagree with the suggestion that evidence obtained in non-Thorndikian situations is irrelevant [here reference is made to the statement in my original note which is quoted at the beginning of this paper] (2, pp. 344-345).

Instead of suggesting that evidence obtained in non-Thorndikian situations was irrelevant to Thorndike's position in *current learning theory*, I quoted some such studies as confirmation of Thorndike's position (3, pp. 153-154).

Postman says he feels that documented criticism should not be construed as a lack of respect for Thorndike and I must confess that this phrasing traps me a little. Is it possible to escape by suggesting that there was no intentional disrespect?

BIBLIOGRAPHY

1. POSTMAN, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489-563.
2. POSTMAN, L. Discussion of Stone's note on the law of effect. *Psychol. Bull.*, 1948, **45**, 344-345.
3. STONE, G. R. A note on Postman's review of the literature on the law of effect. *Psychol. Bull.*, 1948, **45**, 151-160.

BOOK REVIEWS

BEACH, F. A.: *Hormones and behavior. A survey of interrelationships between endocrine secretions and patterns of overt response*. New York: Hoeber, 1948. Pp. xiv+368.

The wide gulf which has long separated physiologists from psychologists is at long last being successfully bridged, at varying points, through the development of new border-line sciences. The present volume of Professor Beach, which describes a field where behavioral psychology and endocrinology merge, is a notable contribution to our understanding of how physiological changes influence psychological events and *vice versa*.

This survey of the effects of hormones upon animal behavior consists of two broad sections. In the first, using a phyletic approach, the author reviews the known hormonal effects upon various types of behavior. While principal attention is given to various aspects of sex behavior (courtship, mating, bisexuality, and parental behavior) there is also a consideration of endocrine influences upon other types of behavior such as migration, aggressiveness, learning and conditioning, general locomotor activity and emotional behavior patterns. For reasons that are not clear, there is an unfortunate omission of many significant clinical findings which should be present in a comprehensive survey of hormone-behavior relationships, *if one includes man*. Thus, no mention is made of the marked psychiatric disturbances associated with endogenous hyperinsulinism, or the effects of insulin upon psychotic behavior, or of the innumerable studies, both positive and negative, of the role of the endocrines in psychosis; and insufficient attention is given to the profound emotional disturbances observed clinically associated with either hyper or hypothyroidism.

The second section of the book is an attempt to bring order out of the conflicting mass of factual data presented in the first section. In this analysis, attention is given to hormonal behavioral responses which operate indirectly via alteration of metabolic or of homeostatic mechanisms; or by control of the morphologic structures involved in specific behavior patterns. The role of external stimulation, whether by light or temperature, or by the presence of other animals, in affecting the function of endocrine glands is considered. Finally, using examples drawn mainly from the field of sex behavior, Beach presents a scholarly analysis of the major sources of variability in hormone-behavioral patterns, and then concludes with a speculative, though intensely interesting, effort to interpret the hormonal effects observed.

Space does not permit an adequate presentation of Beach's views in this short review. However, it should be emphasized that the basic theories formulated, the questions posed, with their attendant influence

upon future lines of investigation, deserve the attention of all workers in this important expanding field.

Since today's scientists are, for the most part, professionals only in their particular field of specialization, the value of the present volume would have been considerably enhanced had the author indicated briefly the broad theoretic structure, the disciplines, and the technics of both endocrinology and the field of animal behavior. This reviewer as an endocrinologist, would have appreciated some discussion of the psychobiological principles which lead Beach to conclude that the facts cited in his survey demonstrate that differences in the hormone-behavior relationships in man and other species are more frequently quantitative than qualitative. Is this conclusion drawn, for example, from the fact that there exist individual differences in maternal efficiency between rats, and in maternal tendencies in women (p. 241) or that there appear to be preferential mating tendencies both in lower animals and in man (p. 243)? Without entering into an extended discussion of this point, it should be mentioned that students who regard behavioral differences between man and lower forms as qualitatively dissimilar, do so on the basis that the *mechanisms* involved are different and not because the overt responses are superficially dissimilar. Similarly, when Beach asserts (p. 279) that "it seems beyond question that progress toward and explanation of the effects of hormones in other animals will inevitably result in better understanding of similar effects in the human," questions arise and are unanswered as to the *nature* and the *level* of human behavioral problems which will be elucidated by animal experimentation on the lower forms.

The transformation of a reasonable tentative opinion advanced by an authoritative writer into a widely-held theory is not an infrequent phenomenon in science. Since it appears without question that Prof. Beach's volume will become an authoritative work on hormones and behavior, it is necessary to point out that some of the conclusions of the author are suspect, because alternative hypotheses were not considered. A single example will suffice to illustrate this point. In discussing the role of the thyroid upon human intelligence (p. 124) Beach notes the serious effects of uncorrected hypothyroidism and the beneficial results of prompt replacement therapy, and then states that "since cretinism or myxedema involve profound systemic abnormality and widespread metabolic deficiency, there is little need to assign to thyroid secretions a specific and exclusive responsibility for maintaining mental efficiency." Clinicians who have observed no comparable deleterious effects upon intelligence with diseases which also involve generalized metabolic deficiency and systemic abnormality (such as untreated juvenile diabetes and Addison's disease) and who have administered dinitrophenol to cretins without influencing their mental status (although repairing, in part their metabolic deficiency), would wonder whether their ob-

servations do not establish some degree of *specificity* for thyroid hormone, as regards intelligence. To be sure, the effect of thyroid secretion upon intelligence may well be indirect and mediated via some chain of metabolic events. To say this, however, is not to deny the possible specificity of thyroid hormone upon those events.

The criticisms mentioned above are reduced to minor significance when the positive features of Prof. Beach's book are considered simultaneously. As the most comprehensive statement of the inter-relationships between the endocrine system and behavior, particularly in lower animals, and as a book where the primary mechanisms of these relationships are lucidly and skillfully discussed, this book will be a source of constant reference to those interested in the psychobiologic principles of behavior.

OSCAR HECHTER.

Worcester Foundation for Experimental Biology.

BURTON, ARTHUR, AND HARRIS, ROBERT E. *Case histories in clinical and abnormal psychology*. New York: Harper, 1947. Pp. xii+680.

This book begins with an introductory piece by Henry A. Murray, the finest writer and most literate man of our field. . . His article should be read for pleasure. It is also an authoritative review of the book, granting that it was Dr. Murray's task to put the book's best foot forward.

In view of the massive nature of the work (669 pages of text), my impressions are best expressed by a series of numbered notes:

1. I had hoped to find here a book of case histories which would be useful to the teacher of an undergraduate course in Abnormal Psychology. What I would have liked is a book of illuminating case histories to supplement the systematic statements of a good textbook. This Burton and Harris do not do. The histories show little gain in insight and organization over the statements found in the usual text. The test material reported is highly technical and could only be explained to a class of first-rate psycho-technicians. Any "Abnormal" teacher who uses the material and who is not intimately at home with the administration and interpretation of test materials cannot avoid being stumped by his class. I believe therefore that the book should be withheld from undergraduate students or beginners in Clinical Psychology.

2. I can cordially commend the book as a *research document* for clinical psychologists and their very advanced students. It shows the formidable body of technique and doctrine which has grown up through the use of various test procedures and is decisive evidence of the "coming of age" of Clinical Psychology.

3. The life-history information in almost all of the cases is *borrowed* from psychiatrists, the clinical psychologist standing by to show the additional data provided by his methods. This is a grave defect, since the reliability and validity of the basic psychiatric information is unknown. In fact, the clinical psychologist appears here as the banner-bearer of good method, the entry point

of behavior science into the field of Psychiatry. The clinical psychologist knows "something" in a positive, comparative, measurable sense. Too often, however, it seems that the silken ears of science, as demonstrated by the clinical psychologist, are sewed on to the complacent sow of Kraepelinian Psychiatry, the whole object not making much sense either in fact or figure of speech. The clinical psychologist must add the instrument of the interview to his arsenal of techniques if he is to give a convincing total account of a case.

4. If one can take this book as evidence, American Clinical Psychology has adjusted itself to a somewhat old-fashioned, "state hospital" type of Psychiatry. I am of the opinion that it will do much better when it must confront the newer, dynamic trends now gaining headway.

5. I don't mean to be too harsh with the old-time psychiatrists. They have been compelled to attempt to make a science out of the scraps of information which are left over when the most severe defense mechanisms of the personality have been called into play. In attempting to build this science, they have fallen back on their biological concepts, have invented a loose set of terms and acquired a set of tricks. It is to this scientific hodge-podge that Clinical Psychology has had to attach itself.

6. The dynamic orientation so deadly necessary in such a book is almost entirely missing. The fitful lightning of psychoanalytic perception is badly needed—that white light which shows the mental case as a human beast trapped and writhing in the grip of culture.

7. The importance of social structural factors is in general neglected, even though class placement is usually noted. The difficulty is that class orientation, class insecurity, mobility, and ethnic membership are seen as a mere "setting" which has to be mentioned but is not perceived as dynamically integrated with mental life itself. Class insecurity can be a vital factor in the system of tensions which conspire to produce a mental disorder.

8. There is a notable absence of a general *theory* of personality development, of habit growth and formation, which makes the case material seem chaotic. The vital role of early acquired, strong, unconscious habits is for all practical purposes belittled and neglected.

9. Much use is made of the Rorschach as if it were already a reliable "test." I am of the opinion that if Rorschach's simple categories turn out to be validly related to mental disorders, Rorschach is going to be known as the luckiest man in the history of science.

10. I see with pain that this adverse review gives me a chance to make forty-five new enemies instead of the usual possible one or two. I hope it will be carefully noted therefore that I do not criticize the *psychological* work or technique of these excellent scientists but only the notion that their book constitutes a generally valuable set of case histories.

In sum then, I would say that this book is particularly useful for clinical psychologists themselves as a reference and research work. It could be helpful to advanced clinical students. It is not, however, the much-needed book of illuminating case histories for the teacher of Abnormal Psychology.

JOHN DOLLARD.

Yale University.

CARROLL, HERBERT A. *Mental hygiene*. New York: Prentice-Hall, 1947. Pp. v+329.

Carroll's "Mental Hygiene" treats material which could be equally well given the title "Abnormal Psychology" or "Mental Hygiene for College Students." The text is organized into fourteen chapters of which four describe and classify the usual abnormal disorders. This group of chapters, forming the main core of the book, is preceded by four chapters dealing respectively with needs for emotional security, mastery, status, and physical satisfactions. It is followed by four chapters which integrate previously presented content with topics such as the school and the community, mental superiority and deficiency, measurement, and regaining mental health.

Carroll has done an excellent job of presenting this material in a direct, readable, interesting style. He has very cleverly incorporated a substantial number of quotations from other sources and references to the literature, without seeming to clutter up his text. Perhaps more important from the point of view of presentation is the fact that he has also accomplished this without detracting from the readability of his style.

He draws heavily upon clinical cases to illustrate his points, the cases usually dealing with the psychological problems of school children or college students. This, of course, has the advantage of presenting material which is of interest to college students, particularly those who are preparing for a teaching career. Other than this case material there is very little in the way of illustrative material with the possible exception of a few tables of data on hospital admission of psychiatric patients.

Carroll has designed this book as an elementary text which his Preface describes as "written with the needs and backgrounds of two groups of students constantly in mind: (1) those who are beginning their work as majors in psychology; (2) those who are not majoring in psychology but are interested in achieving some insights into the dynamics of adjustment which will be of value to them personally and professionally." Since these groups usually have had little or no training in psychology, he has provided general introductory material on motivation, individual differences, learning, and psychometrics.

The book is a relatively small volume which an instructor might want to use as supplementary reading or if he uses this as the main text he may wish to supplement it with another. It contains a few of the inevitable typographical errors but so far as this reviewer observed they in no way detract from the utility of the text. It is regrettable that Prentice-Hall did not see fit to publish this worthwhile book on a better grade of paper. It deserves better treatment than the publisher gave it.

PAUL S. BURNHAM.

Yale University.

BOWERMAN, W. G. *Studies in genius*. New York: Philosophical Library, 1947. Pp. 343.

This is a book about famous people, fame being defined in a particular way. For the first part of the book, which deals with famous Americans, the basic criterion for inclusion was that the individual's biography in the Dictionary of American Biography must extend to $1\frac{1}{2}$ pages. The basic list was modified by excluding some individuals whose claim to fame appeared to the author to stem from notoriety or luck, those who had spent less than half their lifetime in this country, those still living at the time the Dictionary was published, and the like. The list was extended to 1000 names by including some who had slightly less than the specified $1\frac{1}{2}$ pages. Similarly, in the second part, dealing with famous people throughout the world, the basic requirement for inclusion was a biography of at least half a page in the Encyclopedia Britannica. This list was whittled down by excluding a number whose fame seemed to arise primarily from the accident of birth or other reasons not related to their ability or achievement. It is in the above sense that the term "genius" is used in the present work.

Starting with a group defined as above, the author has undertaken to compile a variety of facts about them. These are summarized in tables and presented at some length in the text, with many citations of name and specific fact. The following topics, corresponding to chapters of the text, suggest the types of material covered: place of origin, occupations, heredity and parentage, childhood and youth, marriage and the family, duration of life, wars and epidemics, pathology, height and weight, pigmentation. In most cases, control statistics for the general population could not be presented for the attributes which were studied in this selected group, though the author makes one or two attempts to develop such figures. As a result, the materials provide primarily descriptions of sociological and biological facts about the defined group, rather than critical tests of any hypotheses as to how the members differ from the generality of their contemporaries. In addition to the descriptive statistics and nose-counting, the author provides a certain amount of speculative discussion.

This study of biographical material, written in the tradition of Havelock Ellis' "Study of British Genius" (1904, 1926) and of Cattell's "A Statistical Study of Eminent Men" (1903) may have some interest to psychologists concerned with persons of outstanding achievement. However the treatment does not appear to do much by way of establishing causal relationships or suggesting psychological insights, and therefore its contribution to our understanding of the psychology of "genius" seems quite limited.

ROBERT L. THORNDIKE.

Teachers College, Columbia University.

MILLER, G. A., WIENER, F. M., AND STEVENS, S. S. *Transmission and reception of sounds under combat conditions*. Summary Technical Report of Division 17, NDRC, Volume III. Washington, D. C.* Office of Scientific Research and Development, 1946. Pp. xi+296.

During the recent war the field of military communications was successfully invaded by scientists from many different areas. It was the recognition of the speaking and hearing organism as an integral part of the communications system that brought about a juxtaposition of the psychologist, the physicist, the speech expert and others in a few laboratories whose task it was to investigate basic and developmental problems in speech, hearing, and the processes which intervene between the two. A general acquaintance with the results of this work has been delayed because of the fact that the work has been reported in a large number of individual publications, some of which were classified or were published only by government agencies. For a long time, interested persons have had to content themselves with writing to the Department of Commerce for single reports and have at best got an incomplete picture piecemeal.

This situation has been partially corrected by the publication of a summary report covering the wartime research of the Psycho-Acoustic and Electro-Acoustic Laboratories. The contributions of these laboratories, and to a minor extent of other institutions, have been brought together in a complete, well boiled-down volume called "Combat Instrumentation-II" on the binding and "Transmission and Reception of Sounds under Combat Conditions" on the title-page. Actually both titles are misleading, for in this volume are to be found the basic items on acoustic control and measurement, speech, hearing, factors affecting the intelligibility of speech, effects of amplitude and frequency distortion, psychological and physical characteristics of hearing aids, measurement and development of components of communication systems, etc. Here, indeed, is a handbook which provides direct access to experimental results in this field.

This volume treats two general areas. On the one hand, information of a physical nature is provided concerning sound control, interphone equipment, communications systems in general, radio receivers, hearing aids, and other gadgets. On the other hand, the chapters which are of more immediate interest to psychologists concern some of the more basic functions relating to the ability of the human organism to produce

* Distribution of the Summary Technical Report of NDRC has been made by the War and Navy Departments. Inquiries concerning the availability and distribution of the Summary Technical Report volumes and microfilmed and other reference material should be addressed to the War Department Library, Room 1A-522, The Pentagon, Washington 25, D. C., or to the Office of Naval Research, Navy Department, Attention: Reports and Documents Section, Washington 25, D. C.

and perceive the sounds of speech. Most of the material is presented in graphic form and is explained fully in the accompanying text. The book does not read like a novel; it is highly technical and sometimes fairly difficult. In some respects it is incorrect to speak of the "book" because it seems rather to be a collection of review articles on related fields which are indicated by the chapter headings.

An introductory chapter written by S. S. Stevens orients the reader within the very complicated framework of the OSRD and NRDC research projects. Dr. Stevens also successfully justifies the place of the psychologist in communications research. The following eighteen technical chapters were written by Drs. G. A. Miller and F. M. Wiener** of the Psycho-Acoustic Laboratory. Miller, the psychologist, and Wiener, the communications engineer, combined their efforts to produce a surprisingly homogeneous set of chapters. Theirs was the very difficult task of taking a great number of the individual reports referred to above, sorting them out, combining results and presenting a complete review of these specific areas.

The second chapter on sound control takes up three general matters. First the problem of measuring noise is presented with examples given of the application of measurement of the intensity of narrow bands of noise to specific noisy situations such as the interior of an airplane. Second, the problem of sound control mostly by means of the insertion of sound absorbing materials is presented. Here we find a description of the construction of the large anechoic chamber at Harvard. Finally, the effects of noise on various kinds of human behavior, such as psycho-motor efficiency, are given.

Chapters three and four deal with fundamental characteristics of human hearing and human speech. These are of special interest to psychologists. Here we find graphically presented data on the sensitivity of the ear, the effect of masking on auditory thresholds, the spectra of speech, factors which affect the recognition and intelligibility of speech sounds, and other basic items which would normally have their place in a handbook of experimental psychology. Chapter five presents the development of articulation testing methods. Much of the early research at the laboratories was directed toward establishing some means of evaluating communications systems on the basis of the amount of speech which could be understood. The development of auditory tests whose material ranged from nonsense syllables to complete sentences is given a thorough going over. This information certainly has more than military application. Chapters six, seven and eight analyze the intelligibility of speech. Quantitative data are given which relate intelligibility of speech to different kinds of amplitude and frequency distortions as well as various types of interference.

Chapters nine to fourteen then apply the basic information that is given in chapters three to eight to specific components of communication systems. The interphone seems to be a convenient model for communications systems since it involves the announcer, the microphone, an electronic amplifier, an electro-acoustic transducer, and finally a human ear. The development of components of military communications systems is given in these chapters. We find here the rationale for the development of some of the equipment which is

** Dr. Wiener is now at the Bell Telephone Laboratories.

now familiar to persons in these fields such as the Ear Warden, the Harvintip, different kinds of earphone cushions, noise shields, the use of microphones in gas masks and oxygen masks and other such specific and technical components.

The last five chapters present the results of special projects conducted at the Psycho-Acoustic and Electro-Acoustic Laboratories. First to be considered is the research on hearing aids. In this project commercial hearing aids were evaluated from a physical point of view through the use of equipment that was especially designed for the purpose and also from a psychological point of view with the help of a carefully selected sample of hard-of-hearing persons. Some of the results of this project are already in published form and have found application in the field of audiology. Sonic devices for positioning, direction finding, and indication of true air speed are described in the following two chapters. Another interesting project, the results of which are not too conclusive, had to do with "fly-bar," flying by auditory reference. The results of experimentation on this method of "blind flying" are described. Finally, consideration is given to special devices for use on shipboard. These include radio repeat units, time indication of voice recordings, and lighting of plotting and display surfaces.

In summary, this is a volume of basic information on the characteristics of human speech and human hearing along with some very pertinent physical data on sound control and characteristics of communications equipment. To the reviewer's knowledge there has been no single book containing the kind of information which is presented here on speech and hearing since the publication in 1929 of Harvey Fletcher's "Speech and Hearing." An unqualified recommendation of the book could be made without hesitation except for the fact that too few copies have been published to make it as generally useful as it could be.

IRA J. HIRSH.

Psycho-Acoustic Laboratory, Harvard University.

ADKINS, DOROTHY C. *Construction and analysis of achievement tests.* Washington: U. S. Government Printing Office, 1947. Pp. xvii + 292.

According to the author "the original purpose of this book was to serve as a basis for training personnel of the United States Civil Service Commission and of the United States War Department who were directly engaged in the preparation of written or performance tests of achievement for predicting job performance of public personnel. It was the intention to present basic concepts and methods, not details of operational procedures peculiar to one or both of these particular agencies." The material in the book seems admirably suited for such a training course and has actually been used for this purpose. The orientation of the book is public personnel administration; those interested in educational achievement testing may, therefore, not find the book quite as useful as would be the case if the book were written with their particular problems in mind. There is no material, for example, dealing

with the question as to how one should attempt to measure various outcomes of instruction. The emphasis, instead, is upon the technical matters pertaining to the planning, construction and analysis of achievement tests which are to be used in predicting success on the job.

Chapter I is concerned with the planning of a written test. Topics of special interest in this chapter are: (1) collaboration of subject-matter specialists and test technicians, (2) the nature of job analysis for testing purposes together with a suggested form for making such job analyses, (3) definition of aptitude and achievement tests. It seems to this reviewer that the author might have stressed a bit more that the differences between so-called aptitude and achievement tests are often not very great. An aptitude test is really an achievement test but typically measures knowledge of a more generalized character than that which is found in so-called achievement tests. It would seem more logical to differentiate between aptitude and achievement tests on the basis of purpose rather than content.

The topic of Chapter II is "Constructing and Compiling Written Tests." The discussion of how to convert ideas into multiple-choice test items and determining the kinds of problems which can be set by multiple-choice items is one of the best that this reviewer has seen. The examples of good and bad items illustrate the author's points very well and should be helpful to anyone who prepares tests using the multiple-choice type of item. The author's rich experience in the construction of objective tests is clearly demonstrated in this chapter. The sections on mechanics of recording and preserving items and on the compiling of a written test should be of special interest to all those connected with large-scale testing programs.

Chapter III is entitled "Basic Statistical Tools." The material in this chapter is concerned with those topics in statistics which the test builder should know about, beginning with the frequency distribution and concluding with partial and multiple correlation, problems of sampling and the standard error concept. The reader who is well versed in elementary statistics will not find anything particularly new in this chapter, but the person who wants a brief review of basic statistical tools and the person conducting a training course for test constructors will find the contents of this chapter very helpful.

Chapter IV, the "Analysis of Test Results," is a "must" for anyone—except the most sophisticated—who is concerned with the building of tests. All of the sections in this chapter are treated very competently, but this reviewer liked especially the discussion of (1) the concept of a standardized test, (2) analyzing the difficulty of a test, (3) establishing the validity of a test, (4) weighting the parts of a test or of a total examination, and (5) establishing critical scores and transmuting raw scores. In these sections the author presents many ideas and suggestions which, while they may have been known to others, have not been set down so

they could be used satisfactorily in training courses. The material is excellently organized and clearly presented.

Chapter V, "Special Problems in the Development of Performance Tests," was prepared as a unit independent of the remainder of the book. The material should be of interest to those who wish to prepare better performance tests in the trades fields and various vocational courses. If time and circumstances had permitted, it might have been better to have incorporated this material as a part of Chapter II. The glossary of terms in the appendix is very good and should be especially helpful in instructional work.

In the opinion of this reviewer the author is to be congratulated upon the work which she has done in the writing of this book. In contrast with most books on tests and measurements, this book gets down to fundamentals and really gives the reader some insight into how a test is actually built. All too frequently authors of books in this field have been content with general descriptions of tests and vague generalizations as to how they were constructed. This book will, therefore, be welcomed by all those who are concerned with the fundamentals or foundations of tests and measurements as contrasted with the overviews found so frequently in textbooks published to date.

DEWEY B. STUIT.

State University of Iowa.

GUTHRIE, E. R., & HORTON, G. P. *Cats in a puzzle box*. New York: Rinehart, 1946. Pp. 67.

The problem of this careful study is described as this: "Does the behavior of the cat in the puzzle box go at any point contrary to or in violation of the principle of association?" The authors distinguish, appropriately, between *acts* and the *movements* of which the former are constituted and take "the position that acts are made up of movements that result from muscular contraction, and that it is these muscular contractions that are directly predicted by the principle of association" (p. 7). (Cf. the reviewer's distinction between acts and responses, *Brit. J. Psychol.*, 1931, vol. 22, pp. 150-178.) Again, "The object of the present study is to determine whether the behavior of the cat in the puzzle box indicates that the bare fact of association is an adequate ground for predicting that the associated stimulus cue will be followed by the associated response" (p. 2). Through an argument the cogency of which is not entirely clear, the objective behavior used to indicate the answer to the above questions becomes "the extent to which the cat uses . . . stereotyped movement which would be meaningful only in the light of the past accidents of learning." Does "meaningful" here mean "related to successful accomplishment?" It must not, since it is precisely such notions that the authors regard as useless. But if this is not the case, how is one move-

ment more "meaningful" than another? It should be noted, in passing, that it is not muscular contractions but movements and postures that are observed by the experimenters and recorded photographically. Moreover, it is not always possible to distinguish between movements (which presumably should be described in terms of centimeters, grams, seconds, radians, ergs, etc.) and acts which are described in terms of the change in the relation between animal and environment they bring about. Or, more precisely, what is, objectively, a single movement in the authors' terms may sometimes also be an *act* as they define it. Thus we find movements described as "successful" (pp. 10, 24, 25, 30, 31, etc.), "useless" (p. 20), "unsuccessful" (p. 26), and "futile" (p. 29).

Later in the monograph the inquiry is still further specified: "Our concern was with several problems: (1) What is the nature of the behavioral changes that occur in the course of the experiment? (2) How does success affect the process? (3) Does the principle of association apply to all changes in response to situation?" This follows immediately upon the italicized proposition that "Success is a quality extrinsic to the process of learning and depends on the animal's environment and the accidents of that environment." Again (p. 37): "The real problem of the puzzle box is not how the successful movement comes about. . . . The real problem is this: how describe the process which appears to eliminate a great many responses and to preserve that specific movement series that has resulted in success?"

The behavior of the cats is reported in two ways: summary verbal description and a complete record of escape postures for 13 cats and one dog, reproduced in silhouettes traced from the negatives for all trials except a few in which the camera failed. This is in general an admirable method of reporting the particular behavioral detail that is the authors' main interest and it enables the reader to form his own judgment of the uniformity or stereotypy of the animal's posture at the instant when the door of the puzzle box opens. Even more illuminating is the excellent film of the same title as the present monograph published by the authors a year or two earlier, which shows one entire sequence of trials (Cat A) and a series following the fortieth trial for Cat K. Incidentally, this latter reveals that final postures that look alike in the silhouettes are not always the same nor arrived at in the same way. Thus Cat K once failed in her usual progress from entrance to exit to lean far enough against the trigger mechanism (a pole near the entrance) to set it off. Then, noting that the door had failed to open (I do not see how even the authors' dogged opposition to cognitive restructuring and the role of success can avoid this interpretation!), she stopped abruptly, *reversed her movement, backed up*, and leaned harder against it! Thus, although the posture at the instant of opening may look the same as on the trials when one push sufficed, it cannot conceivably be muscularly or kinasthetically the same since it is arrived at, so to speak, in reverse.

This incident also yielded evidence of stereotypy: "From that time on *for several trials* [italics mine] the cat paused and made two distinct movements against the pole, though the first was successful" (p. 40). Why only "for several trials?"

The authors summarize their generalizations of the puzzle box behavior as follows (pp. 41-42): "The behavior of the cat on one occasion tends to be repeated on the next, even to occasional prolonged series of movements about the cage. Exceptions to this are either the result of a different entrance, which initiates a different line of action, or the result of accidental distractions, which may deflect the behavior from its former sequence; moreover, when the cat has been in the box for a long time, responses made later in the trial may supplant those made earlier. . . . The present account explains the stability of the final movement of escape through the protection of that response from unlearning, a protection furnished by the fact that escape separates animal and situation and gives no opportunity for learning new responses to the situation." But this protection from unlearning is hardly an explanation of learning.

One sometimes wonders what all the shooting is about. The aim is to show that the results of actions are irrelevant to the learning process and the conclusions imply that this is the case; but throughout there is implicit recognition that they are relevant, even if only as terminals of acts. "Even so simple an act as clawing at the door alters the cat's relation to the door. It is now a cat that has clawed rather than a cat that is about to claw. What its behavior was prepared for [*sic!*] by previous experience did not happen" (p. 40).

Perhaps a reviewer may be permitted to offer an alternative interpretation of these interesting data. This one is brazenly anthropomorphic; Stereotypy of behavior occurs when (1) the behavior is the most economical means (for the animal and situation in question) to the satisfaction of a need and (2) when the "successful movement" is for any reason (e.g., embeddedness in other movements, apparent irrelevance as in the present experiments, etc.) not identifiable. This latter case is related to the ritualistic acts developed by athletes, coaches, anglers, hunters, concert artists, etc. The authors' vast improvement on the puzzle box method offers a ready means for varying the apparent nexus between act and result and thus of testing this hypothesis.

DONALD K. ADAMS.

Duke University.

INDEX OF SUBJECTS

- Absolute pitch
 - note on Neu's review, 161
 - reply to Bachem, 534
- Acuity, physiological basis of visual, 465
- Addends, use in experimental control, social census and managerial research, 41
- Analysis of the use of the interruption technique in experimental studies of "repression," 491
- Analysis of variance—repeated measurements, 131
- Appetite, palatability and feeding habit—a critical review, 289
- Attitudes, techniques for the diagnosis and measurement of intergroup, 248

- Bachem, reply to, 534
- Behavior, influence of work on, 1
- Binaural summation—a century of investigation, 193
- Bureau of Naval Personnel, test development and personnel research in, 144

- Color terms and definitions, 207
- Corrigenda, 345
- Current trends in psychology, 75

- Definitions, color terms, 207
- Design and analysis of psychological experiments, latin square principle in, 427
- Discussion of Stone's note on the law of effect, 344
 - reply to, 536
- Effect, law of
 - discussion of Stone's note on, 344
 - note on Postman's review of, 151
- Evaluation of the study of Bernardine G. Schmidt entitled "Changes in the social and emotional behavior of children originally classified as feeble-minded," 321
 - reply to, 334
- Experimental control, use of addends in, 41
- Experimental studies of "repression,"
 - analysis of use of interruption technique in, 491
- Eysenck, suggestibility and narcosis—a reply to, 346
- Feeble-minded children, evaluation of Bernardine G. Schmidt's study of, 321
 - reply to, 334
- Factor analysis, technic of homogeneous tests compared to, 507
- Homogeneous tests compared to scale analysis and factor analysis, 507
- Influence of work on behavior, 1
- Intergroup attitudes and behavior, techniques for diagnosis and measurement of, 248
- Interruption-technique, use in experimental studies of "repression," 491
- Inventories (personality), validity in military practice, 385
- Kinsey's "Sexual behavior in the human male": some comments and criticisms, 443
- Latin square principle in the design and analysis of psychological experiments, 427
- Law of effect
 - discussion of Stone's note on, 344
 - note on Postman's review of, 151
- Managerial research, use of addends in, 41
- Measurement of
 - intergroup attitudes and behavior, 248
 - transfer of training, 97
- Military practice, validity of personality inventories in, 385
- Narcosis
 - rejoinder on suggestibility and, 163
 - reply to Eysenck on suggestibility and, 346

- Neu's review on absolute pitch, note on, 161
 reply to, 534
 Note on Neu's review of the literature on absolute pitch, 161
 reply to, 534
 Note on Postman's review of the literature on the law of effect, 151
 discussion of, 344

 Object sampling, subject and, 530
 Olfactometry, techniques in, 231

 Personality inventories, validity in military practice, 385
 Personnel research and test development in the Bureau of Naval Personnel, 144
 Pitch, absolute
 note on Neu's review, 161
 reply to Bachem, 534
 Postman, reply to, 536
 Postman's review on the law of effect, note on, 151
 discussion of, 344
 Physiological basis of visual acuity, 465
 Psychological experiments, latin square principle in, 427
 Psychology, current trends in, 75

 Rejoinder, (on) suggestibility and narcosis, 163
 reply to, 346
 Repeated measurements, analysis of variance in, 131
 Reply
 to Bachem, 534
 to Postman, 536
 (to Kirk's evaluation of Schmidt's study), 334
 Repression, analysis of use of interruption technique in study of, 491

 Sampling, subject and object, 530

 Scale analysis, technic of homogeneous tests compared to, 507
 Schmidt, Bernardine G., evaluation of study of, 321
 reply to, 334
 "Sexual behavior in the human male," comments and criticisms, 443
 Social census, use of addends in, 41
 Stone's note on the law of effect, discussion of, 344
 Study by Bernardine G. Schmidt, evaluation of, 321
 reply to, 334
 Subject and object sampling—A note, 530
 Suggestibility and narcosis—
 a rejoinder, 163
 a reply to Eysenck, 346
 Summation, binaural, 193

 Technic of homogeneous tests compared with certain aspects of scale analysis and factor analysis, 507
 Techniques for the diagnosis and measurement of intergroup attitudes and behavior, 248
 Techniques in olfactometry—a critical review of the last hundred years, 231
 Test development, in the Bureau of Naval Personnel, 144
 Training, measurement of transfer of, 97
 Transfer of training, measurement of, 97
 Trends, in psychology, 75

 Use of addends in experimental control, social census and managerial research, 41

 Validity of personality inventories in military practice, 385
 Variance, analysis of, 131
 Visual acuity, physiological basis of, 465

 Work, influence on behavior, 1

INDEX OF AUTHORS

ORIGINAL CONTRIBUTIONS, SHORT ARTICLES, SPECIAL REVIEWS, REPORTS, NOTES

- | | |
|----------------------|------------------------|
| Bachem, A., 161 | Johnson, H. M., 345 |
| Boring, E. G., 75 | Kirk, S. A., 321 |
| Brennan, J. G., 207 | Kogan, L. S., 131 |
| Burnham, R. W., 207 | Loevinger, J., 507 |
| Conrad, H. S., 385 | Neu, D. M., 534 |
| Crowley, M. E., 97 | Newhall, S. M., 207 |
| Deri, S., 248 | Pepitone, A. D., 248 |
| Dinnerstein, D., 248 | Postman, L., 344 |
| Ellis, A., 385 | Schmidt, B. G., 334 |
| Eysenck, H. J., 163 | Senders, V. L., 465 |
| Flanagan, J. C., 144 | Snyder, W. U., 346 |
| Foster, H., 97 | Solomon, R. L., 1 |
| Gagne, R. M., 97 | Stone, G. R., 151, 536 |
| Glixman, A. F., 491 | Terman, L. M., 443 |
| Grant, D. A., 427 | Toops, H. A., 41 |
| Hammond, K. R., 530 | Wenzel, B. M., 231 |
| Harding, J., 248 | Young, P. T., 289 |
| Hirsh, I. J., 193 | |

BOOKS REVIEWED

- | | |
|----------------------------|----------------------|
| Adkins, D. C., 546 | Gemelli, A., 360 |
| Allport, G. W., 171 | Giese, A. C., 358 |
| Axline, V. M., 281 | Guthrie, E. R., 548 |
| Bartley, S. H., 355 | Hall, R. B., 177 |
| Baumgarten-Tramer, F., 183 | Hall, V. E., 358 |
| Beach, F. A., 538 | Hamilton, G., 367 |
| Beck, H. P., 92 | Harms, E., 369 |
| Bowerman, W. G., 543 | Harrower, M. R., 275 |
| Burnham, P. S., 184 | Harris, R. E., 540 |
| Burton, A., 540 | Hartwell, S. W., 376 |
| Cameron, N., 372 | Jersild, A. T., 281 |
| Cantril, H., 165 | Joad, C. E. M., 288 |
| Carmichael, L., 357 | Kinsey, A. C., 272 |
| Carroll, H. A., 542 | Kitay, P. M., 178 |
| Cleeton, G. U., 95 | Kluckhohn, C., 91 |
| Chute, E., 355 | Köhler, W., 351 |
| Cole, L., 366 | Horton, G. P., 548 |
| Crawford, A. B., 184 | Kitchen, P. C., 378 |
| Crismon, J. M., 358 | Kracauer, S., 173 |
| Dearborn, W. F., 357 | Kraft, M. A., 95 |
| Erickson, C. E., 186 | Le Cron, L. M., 283 |
| Frederick, R. W., 378 | Leighton, D., 91 |
| Garrett, H. E., 87 | McFarland, R. A., 93 |

Marks, R. W., 377
 Marquis, D. G., 89
 Mase, D. J., 277
 McElwee, A. R., 378
 Merrill, M. A., 279
 Miller, G. A., 544
 Morgan, J. J. B., 366
 Murphy, G., 348
 Mursell, J. L., 181
 Nielson, J. M., 374
 Oden, M. H., 363
 OSS Assessment Staff, 460
 Peatman, J. G., 380
 Postman, L., 171
 Richards, T. W., 275
 Robinson, F. P., 378
 Royster, R. F., 95
 Sadler, W. S., 278

Sherif, M., 165
 Sherman, M., 287
 Smith, G. M., 181
 Snyder, W. U., 371
 Sorokin, P., 175
 Stevens, S. S., 544
 Terman, L. M., 363
 Thompson, G. N., 374
 Thurstone, L. L., 85
 Tiffin, J., 285
 Tomkins, S. S., 361
 Tredgold, A. F., 376
 Wiener, F. M., 544
 Williams, R. M., 361
 Wolff, W., 359
 Woodworth, R. S., 89
 Wortis, S. B., 179
 Zunini, G., 286, 360

BOOK REVIEWERS

Adams, D. K., 548
 Anastasi, A., 366
 Asch, S. E., 171
 Ball, J., 287
 Bartley, S. H., 358
 Beach, F. A., 181
 Benton, A. L., 376
 Bergmann, G., 355
 Blair, G. M., 380
 Brown, J. F., 374
 Burnham, P. S., 542
 Corsini, R., 287
 Darley, J. G., 185
 Dollard, J., 540
 Dunlap, J. W., 95
 Fearing, F., 175
 Fischer, R. P., 375
 Fite, M. D., 369
 Frenkel-Brunswik, E., 351
 Gardner, J. W., 178
 Graham, J. L., 179
 Grant, D. A., 89, 181
 Gregory, W. S., 372
 Gundlach, R. H., 177
 Harlow, H. F., 359
 Harris, D. B., 281
 Hartmann, G. W., 360
 Hartstein, J. I., 281
 Hechter, O., 538
 Hewson, L. R., 377

Hirsh, I. J., 544
 Horst, P., 87
 Hunt, W. A., 275, 461
 Jenness, A., 173
 Kappauf, W. E., 96
 Klineberg, O., 363
 Landis, C., 275
 Loevinger, J., 381
 McGehee, W., 286
 Magaret, A., 277
 Marcuse, F. L., 284
 Metcalf, J. T., 361
 Meyerson, L., 278
 Peak, H., 288
 Riess, B. F., 93
 Roberts, S. O., 367
 Rose, A. A., 184
 Sappenfield, B. R., 183
 Sargent, H., 462
 Schlosberg, H., 90
 Sears, P. S., 283
 Snyder, W. U., 279
 Stuit, D. B., 546
 Super, D. E., 188
 Thorndike, R. L., 543
 Welch, L., 370
 Worchel, P., 357
 Young, K., 92
 Young, P. C., 378

Reprinted:

Opinion-Attitude Methodology

By

Quinn McNemar

This popular issue has been reprinted so that it is again available.

It presents an appraisal of the techniques and methods used in opinion and attitude research.

Psychological Bulletin, July, 1946

No. 4

85 pages

\$1.25

American Psychological Association

1515 Massachusetts Avenue N.W.

Washington 5, D.C.

STUDIES IN PSYCHOSOMATIC MEDICINE

**An Approach to the Cause and Treatment
of Vegetative Disturbances**

By **FRANZ ALEXANDER** and **THOMAS M. FRENCH**, both of the
Chicago Institute for Psychoanalysis, and 18 Contributing Specialists

A new book consisting of a collection of papers, based on psychoanalytic study of patients suffering from chronic disturbances of the vegetative organs, prepared by the Staff of the Chicago Institute for Psychoanalysis over the past 16 years. The more significant papers are here presented as a method of approach to the analysis of persons using all available information. Of especial interest to all seeking to understand how body and mind work together to produce such disturbances. \$7.50

PSYCHOLOGY AND ETHICS

A Study of the Sense of Obligation

By **HARRY L. HOLLINGWORTH**, *Professor Emeritus of Psychology, Columbia University*

A psychologist's explanation of conduct and moral principles in terms of motivation, learning, and control, based on his scientific inspection of ethical topics. In addition to his scientific interest in ethical problems, the author believes his conclusions may provide the reader with the opportunity and encouragement to consider a design for living. The book will prove useful in courses in education, social psychology, human relations, and ethics. \$3.50

THE ABNORMAL PERSONALITY

By **ROBERT W. WHITE**, *Director of the Psychological Clinic, Lecturer in Clinical Psychology, Harvard University*

A text for the introductory course in abnormal psychology. Its theme is disordered personalities: people who are maladjusted, neurotic, delinquent, psychotic, brain-injured, or in some other way disturbed in their personal reactions to life. Of the first two chapters, one is historical, the other clinical. The author has then selected five case histories representing a wide range of disorders to serve as examples of illustrating problems and principles of abnormal psychology. These are referred to throughout the book. \$5.00

THE RONALD PRESS COMPANY

75 East 26th Street

